YURI ROZANOV

PROCESOS ALEATORIOS

curso resumido

EDITORIAL MIR-MOSCU

ИЗДАТЕЛЬСТВО
МИР

Ю. А. РОЗАНОВ

СЛУЧАЙНЫЕ ПРОЦЕССЫ

ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА

YU. A. ROZANOV

PROCESOS ALEATORIOS

curso resumido

EDITORIAL MIR MOSCU

CDU 519.2=60

Traducción del ruso

Impreso en la URSS
1973

На испанском языке

© Traducción al espanol. Editorial Mir. 1973

P $\frac{0223-367}{041\,(01)-73}$

INDICE

Del autor 7

CAPITULO I

INTRODUCCIÓN A LOS CONCEPTOS FUNDAMENTALES DE LA TEORÍA DE PROBABILIDADES

§ 1. Espacio de sucesos elementales y probabilidad 9
1. Prueba con resultados equiprobables (9). 2. Espacio de los sucesos elementales (15). 3. Propiedades fundamentales de la probabilidad. Aditividad y continuidad (19). 4. Concordancia entre el modelo y la prueba (23).

§ 2. Independencia y probabilidad condicional 25
1. Concepto de independencia (25). 2. Probabilidad condicional (29).

§ 3. Magnitudes aleatorias y distribución de las probabilidades. Independencia 33
1. Distribuciones discreta y continua (33). 2. Distribución conjunta de las probabilidades (35). 3. Transformaciones de las magnitudes aleatorias (39). 4. Distribuciones condicionales de probabilidades (42). 5. Magnitudes aleatorias multidimensionales (44).

§ 4. Esperanzas matemáticas de las magnitudes aleatorias 45
1. Esperanza matemática, definición y algunas fórmulas (45). 2. Momentos, dispersión y desigualdad de Chébishev (51). 3. Esperanza matemática condicional (53). 4. Distancia media cuadrática y coeficiente de correlación (57). 5. Algunos teoremas sobre la convergencia (61). . . . . .

§ 5. Series infinitas de pruebas independientes y ley de los grandes números . . . . . .
1. Ley de los grandes números (67). 2. Probabilidad y frecuencia (70).

CAPITULO II

ALGUNAS DISTRIBUCIONES DE PROBABILIDADES

§ 1. Elección aleatoria y variación aleatoria 72
1. Algunas fórmulas combinatorias (72). 2. Distintas distribuciones de las partículas independientes en el espacio fásico (76).

§ 2. Distribución de Poisson. Corriente uniforme de sucesos y tiempo de espera 84
1. Distribución de Poisson en las partículas (84). 2. Tiempo de espera del suceso aleatorio (88).

§ 3. Pruebas de Bernoulli. Movimiento browniano y distribuciones de probabilidades ligadas a éste 95
1. Pruebas de Bernoulli y distribución binomial. Aproximaciones de Poisson y normal (95). 2. Proceso del movimiento browniano. Distribución de probabilidades del máximo y momento de su consecución (100).

§ 4. Distribución normal de probabilidades y distribuciones ligadas a ella 108
1. Distribución normal multidimensional (108). 2. Valuación de

los parámetros de la distribución normal. La distribución χ^2 y la distribución de Student (115).

§ 5. ***Distribución de probabilidades y funciones características*** 121
1. Funciones características y sus propiedades fundamentales (121). 2. Convergencia de las distribuciones de probabilidades (127).

CAPITULO III

ALGUNOS MODELOS DE PROCESOS ALEATORIOS

§ 1. **Algunas definiciones y ejemplos** 135
1. Definición general del proceso aleatorio (135). 2. Los procesos aleatorios de Márkov (136).

§ 2. **Cadenas de Márkov. Clasificación de los estados. Distribuciones estacionarias** 141
1. Probabilidades de paso (141). 2. Estados reversibles e irreversibles (145). 3. Tiempo medio de estancia en un estado. Clasificación de los estados (149). 4. Teorema ergódico (convergencia hacia la distribución estacionaria) (153).

§ 3. ***Cadenas de Márkov (tiempo continuo)*** 161
1. Ecuaciones diferenciales para las probabilidades de paso (161). 2. Coeficiente de ergodicidad y convergencia hacia la distribución estacionaria (167).

§ 4. **Procesos que se ramifican** 170
1. Ecuación diferencial de las funciones generatrices (170). 2. Efectos de degeneración y explosión (176).

§ 5. **Algunos procesos de servicios en masa y fluctuaciones aleatorias** 177
1. Procesos de restablecimiento (177). 2. Sucesión de sumas de las magnitudes independientes. Distribución del máximo (183). 3. *Procesos aleatorios en los sistemas con una línea de servicio (191).*

§ 6. **Procesos aleatorios en los sistemas lineales** 199
1. Algunas observaciones auxiliares (199). 2. Integral estocástica (202). 3. Convergencia hacia el proceso estacionario (207). 4. Procesos de efecto fraccionario (209).

§ 7. **Procesos aleatorios estacionarios** 214
1. Representación espectral de los procesos estacionarios y transformación de Fourier (214). 2. Transformaciones lineales. Ejemplos (224).

§ 8. **Procesos de difusión** 232
1. Procesos aleatorios representados por la integral estocástica Ito (232). 2. Ecuaciones diferenciales de Kolmogórov (245).

CAPITULO IV

ALGUNAS TAREAS DE PRONOSTICACIÓN, FILTRADO Y REGULACIÓN DE LOS PROCESOS ALEATORIOS

§ 1. **Tarea general sobre la aproximación óptima. Ejemplos** 252

§ 2. **Pronosticación y filtrado de los procesos aleatorios estacionarios** 259
1. Tarea de pronosticación lineal (259). 2. Filtrado lineal (valuación del valor medio) (263).

§ 3. ***Esperanzas matemáticas condicionales y algunas tareas de pronosticación y de filtrado*** 271
1. Una vez más sobre las esperansas matemáticas condicionales (271). 2. Papel de las probabilidades a posteriori en algunas tareas de pronosticación y filtrado (276).

Indice temático 283

DEL AUTOR

Este libro ha sido escrito paralelamente al Curso de lecciones sobre la teoría de los procesos aleatorios que expuse en el Instituto físico-técnico de Moscú en los años 1969-1970. Tengo la esperanza de que será útil, en uno u otro grado, para todo el que desee dominar los resultados y métodos fundamentales de esta teoría.

Quisiera expresar aquí mi reconocimiento a Yu. V. Prójorov que leyó el manuscrito del libro e hizo una serie de observaciones críticas útiles, que me obligaron a examinar otra vez, más detalladamente, todo el contenido e introducir muchos cambios. Mi reconocimiento a O. V. Viskov que tomó en sí el trabajo de redacción del libro.

Yu. A. Rózanov

CAPITULO I

INTRODUCCIÓN A LOS CONCEPTOS FUNDAMENTALES DE LA TEORÍA DE PROBABILIDADES

§ 1. ESPACIO DE SUCESOS ELEMENTALES Y PROBABILIDAD

1. Prueba con resultados equiprobables. Empezamos nuestro curso con la descripción de algunas pruebas elementales, en las cuales el resultado no se puede de antemano predeterminar con certidumbre, sino sólo se puede hablar sobre la probabilidad de tal o cual resultado.

Lanzamiento de la moneda. A esta prueba se recurre, corrientemente, al realizar un sorteo, cuando la moneda simétrica se lanza con un fuerte giro y como resultado cae de forma aleatoria «cara» o «cruz» (fig. 1).

Fig. 1.

Lanzamiento del dado de juego. Este ensayo es una parte componente del juego popular infantil conocido por el nombre «Quien va despacio, va más lejos», «Circo», etc., en el cual, después del lanzamiento correspondiente del cubo (dado de juego) regular con las caras numeradas, en dependencia del número de puntos que han caído, el jugador desplaza su figura (según las reglas de juego) en el número correspondiente de pasos (casillas). Durante el lanzamiento del dado de juego cae una u otra cara, con tal o cual número de puntos $a = 1, 2, \ldots, 6$

Algo más compleja es la prueba que se realiza en el conocido juego de azar cuando cada uno de los jugadores lanza por turno dos dados de juego, y gana el que tiene la suma mayor de puntos. Al lanzar dos dados de juego cae de una forma aleatoria una de las combinaciones posibles (a, b), donde a significa el número de puntos que han caído en el primer dado, y b el número de puntos en el segundo dado.

Juego a la ruleta. Imagínese un disco pesado, dividido en N sectores regulares, que se encuentra en posición horizontal y puede girar fácilmente alrededor de su eje; a lo largo de la circunferencia por el borde del disco se tiene un hundimiento (canal) uniforme en el cual se encuentra una bolita pequeña que se desplaza libremente (fig. 2). A cada paso por separado se le comunica al disco una rotación fuerte, con la cual rueda la bolita por el canal. Después de pararse el disco, se para también la bolita, cayendo en uno de los sectores (designados en el disco con los números desde 1 hasta N). En dependencia del número aleatorio que cayó $\nu = 1, \ldots, N$ y de la «puesta hecha» el jugador recibe el premio correspondiente.

Fig. 2.

Con relación a cada una de las pruebas descritas anteriormente (lanzamiento de la moneda o del dado de juego, el «lanzamiento» de la bolita en el juego a la ruleta), se puede decir lo siguiente: primero, el resultado de la prueba es aleatorio; segundo, se tiene un número finito de resultados diferentes que se excluyen mutuamente uno al otro; tercero, todos los resultados son equiprobables.

Cualquier prueba de tal naturaleza la denominaremos *prueba con un número finito de resultados equiprobables*, llamando a estos resultados *sucesos elementales* y designándolos en lo sucesivo por ω.

Designemos por N el número total de sucesos elementales. De acuerdo a que todos estos sucesos son equiprobables, consideraremos que la probabilidad de cada uno de ellos es igual a $1/N$.

Considerando tal o cual suceso A, que se realiza (o no se realiza) en dependencia del resultado aleatorio ω, designemos por $N(A)$ el número total de resultados elementales ω, que conducen a la realización de este suceso. Está claro, que el suceso A será tanto más p r o b a b l e[1]) cuanto mayor sea el número $N(A)$. Consideramos que la probabilidad del suceso A, representándola por $\mathbf{P}(A)$, es proporcional a la magnitud $N(A)$; más exactamente definimos a esta probabilidad así:

$$\mathbf{P}(A) = \frac{N(A)}{N}. \tag{1.1}$$

Por ejemplo, la probabilidad de que caiga «cara» al lanzar la moneda es igual a 1/2; la probabilidad del suceso de que «al lanzar el dado de juego cae un número par de puntos» también es igual a 1/2; al lanzar dos dados de juego la probabilidad del suceso A de que

[1]) La expresión del tipo «el suceso A es m á s p r o b a b l e que el suceso B», que se comprende intuitivamente debe ser interpretada, a nuestro parecer, en el sentido de que, cuando la prueba se repita reiteradamente el suceso A se verificará más f r e c u e n t e m e n t e, que el suceso B.

«la suma de los puntos que cayeron sea mayor de 10» es igual a 1/12, ya que el número de todos los resultados elementales posibles ω, es decir, de combinaciones distintas de puntos (a, b), donde $a, b = 1, \ldots, 6$, es $N = 36$, y el número de resultados, para los cuales se realiza el suceso A (es decir, cuando $a + b > 10$), es $N(A) = 3$, ya que al suceso A sólo llevan los resultados (5, 6), (6, 5), (6, 6).

Volvamos a la prueba con el disco giratorio, sobre el cual se encuentra la bolita (véase «Juego a la ruleta»). Después de pararse el disco la bolita ocupa una posición determinada, que se puede describir por la coordenada angular ω, $0 \leqslant \omega \leqslant 2\pi$, si consideramos la bolita como un punto sobre la circunferencia.

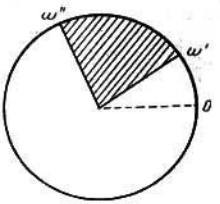

Fig. 3.

El punto aleatorio ω cae en cualquiera de los N sectore siguales, en que está dividido el disco, con una misma probabilidad $1/N$. Considerando que la probabilidad de que caiga en cualquiera de los sectores $A = [\omega', \omega'')$ (fig. 3) es una misma para todos los sectores con la misma distancia angular $\omega'' - \omega'$, determinemos esta probabilidad $\mathbf{P}(A)$ así:

$$\mathbf{P}(A) = \frac{\omega'' - \omega'}{2\pi}.$$

Luego, desarrollando la circunferencia (de longitud L) en el segmento $[0, L]$ de la recta real $-\infty < x < \infty$, sobre el punto $\xi = \frac{\omega}{2\pi} L$ diremos que, en dependencia del caso, cae en tal o cual intervalo $[x', x'')$ de este segmento, y precisamente, cae en cualquier intervalo de longitud $x'' - x'$ con una misma probabilidad igual a $\frac{x'' - x'}{L}$

El valor ξ que es la coordenada del punto considerado depende del caso, hablando de otra forma, es *una magnitud aleatoria*. La probabilidad de que ella tome un valor en el intervalo $[x', x'')$ (designado más adelante por $\mathbf{P}\{x' \leqslant \xi \leqslant x''\}$) se puede determinar por la fórmula siguiente:

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = \int_{x'}^{x''} p_\xi(x)\, dx, \tag{1.2}$$

donde la función subintegral $p_\xi(x)$, llamada en adelante *densidad de probabilidad* para la magnitud casual ξ que consideramos, es

$$p_\xi(x) = \begin{cases} \frac{1}{L} & \text{para } 0 \leqslant x \leqslant L, \\ 0 & \text{para } x < 0,\ x > L. \end{cases} \tag{1.3}$$

En el ejemplo expuesto la magnitud ξ está *distribuida uniformemente* en el segmento $[0, L]$.

Sea $y = \varphi(x)$ una función cualquiera en el segmento $[0, L]$, que tiene una derivada positiva $\varphi'(x)$ continua a trozos; ella transforma biunívocamente el segmento $[0, L]$ en otro determinado segmento $[a, b]$ (que puede incluso ser infinito) (fig. 4). Examinemos la nueva magnitud aleatoria $\eta = \varphi(\xi)$. ¿Qué probabilidad existe de que el valor de η descanse en el intervalo $\{y' \leqslant \eta \leqslant y''\}$?

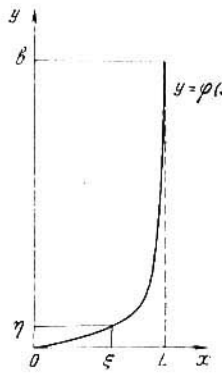

Fig. 4. Se ve fácilmente que para la función $y = \varphi(x)$ de la forma indicada el punto aleatorio $\eta = \varphi(\xi)$ cae con gran probabilidad en la mitad inferior del segmento $[a, b]$, $a = 0$.

Designemos por $x = \psi(y)$ la función en el segmento $[a, b]$, que es inversa de la función inicial $y = \varphi(x)$, $0 \leqslant x \leqslant L$. Evidentemente, los sucesos $\{y' \leqslant \eta \leqslant y''\}$ y $\{\psi(y') \leqslant \xi \leqslant \psi(y'')\}$ coinciden, y la probabilidad del suceso $\{\psi(y') \leqslant \xi \leqslant \psi(y'')\}$, de acuerdo a la fórmula (1.2) es igual a $\int_{\psi(y')}^{\psi(y'')} p_\xi(x)\,dx$. Utilizando el cambio de variable $x = \psi(y)$, obtenemos que la probabilidad buscada del suceso $\{y' \leqslant \eta \leqslant y''\}$ es

$$\mathbf{P}\{y' \leqslant \eta \leqslant y''\} = \int_{\psi(y')}^{\psi(y'')} p_\xi(x)\,dx =$$

$$= \int_{y'}^{y''} p_\xi[\psi(y)]\,\psi'(y)\,dy = \int_{y'}^{y''} p_\eta(y)\,dy,$$

donde la función $p_\eta(y)$ tiene la forma

$$p_\eta(y) = \begin{cases} p_\xi[\psi(y)]\dfrac{1}{\varphi'[\psi(y)]} & \text{para } a \leqslant y \leqslant b, \\ 0 & \text{para } y < a,\ y > b, \end{cases} \tag{1.4}$$

y en nuestro caso $\left(\text{cuando } p_\xi(x) = \frac{1}{L},\ 0 \leqslant x \leqslant L\right)$

$$p_\eta(y) = \frac{1}{L} \cdot \frac{1}{\varphi'[\psi(y)]},\ a \leqslant y \leqslant b.$$

Vemos que la magnitud aleatoria η puede tener *una densidad de probabilidad* muy compleja $p_\eta(y)$, $-\infty < y < \infty$, con ayuda de la cual la probabilidad de caer en tal o cual intervalo $[y', y'']$ se determina como

$$\mathbf{P}\{y' \leqslant \eta \leqslant y''\} = \int_{y'}^{y''} p_\eta(y)\,dy \left(p_\eta(y) \geqslant 0,\ \int_{-\infty}^{\infty} p_\eta(y)\,dy = 1\right)$$

Supongamos ahora que tenemos dos magnitudes aleatorias ξ_1 y ξ_2, análogas a la magnitud aleatoria ξ examinada anteriormente, que no están ligadas entre sí de ningún modo. Digamos, el punto ξ_1 se lanza «al azar» sobre el segmento $[0, L_1]$ del eje $-\infty < x_1 < \infty$, y el punto ξ_2, sobre el segmento $[0, L_2]$ del eje $-\infty < x_2 < \infty$ del sistema de coordenadas cartesiano en el plano.

Considerando a (ξ_1, ξ_2) como un punto aleatorio en el plano con las coordenadas ξ_1 y ξ_2, análogamente a la fórmula (1.2) determinamos la probabilidad de caída de este punto en una u otra zona A (con frontera lisa a trozos) del modo siguiente:

$$\mathbf{P}(A) = \iint_A p_{\xi_1 \xi_2}(x_1, x_2)\, dx_1\, dx_2 \tag{1.5}$$

donde la función subintegral $p_{\xi_1 \xi_2}(x_1, x_2)$ es

$$p_{\xi_1 \xi_2}(x_1, x_2) = \begin{cases} \dfrac{1}{L_1 L_2} \text{ para } 0 \leqslant x_1 \leqslant L_1,\ 0 \leqslant x_2 \leqslant L_2, \\ 0 \text{ para los restantes } x_1,\ x_2. \end{cases} \tag{1.6}$$

Para la zona A del rectángulo $\Omega = \{0 \leqslant x_1 \leqslant L_1,\ 0 \leqslant x_2 \leqslant L_2\}$ donde, solamente, cae el punto (ξ_1, ξ_2) la probabilidad $\mathbf{P}(A)$ de caída en A determinada por la fórmula (1.5) es numéricamente igual a la relación entre el área de la zona A y el área de todo el rectángulo Ω.

Ejemplo. Supongamos que lanzamos al azar sobre el segmento de longitud L dos puntos independientemente uno del otro. ¿Qué probabilidad hay de que la distancia entre ellos no sea mayor que l?

Designemos por ξ_1, ξ_2 las coordenadas de estos puntos en el segmento $[0, L]$. Luego, tomemos ξ_1 en el segmento $(0, L)$ del eje x_1, ξ_2 en el segmento $(0, L)$ del eje x_2 y consideremos que el punto (ξ_1, ξ_2) se lanza al azar en el cuadrado $\Omega = \{0 \leqslant x_1 \leqslant L,\ 0 \leqslant x_2 \leqslant L\}$. Evidentemente, la probabilidad buscada de que $\Delta = |\xi_1 - \xi_2| \leqslant l$, coincide con la probabilidad $\mathbf{P}(A)$ de caída del punto (ξ_1, ξ_2) en la zona A, limitada por las rectas $x_2 = l + x_1$ y $x_2 = -l + x_1$ (véase la fig. 5, donde esta zona ha quedado sin rayado). De tal modo, de acuerdo con la fórmula (1.5) la probabilidad $\mathbf{P}(A)$ puede ser expresada así:

$$\mathbf{P}(A) = \iint_A \frac{1}{L^2}\, dx_1\, dx_2 = \frac{2Ll - l^2}{L^2},$$

donde $2Ll - l^2$ es el área de la zona A.

Ejemplo. (*Tarea de Buffon sobre la aguja*). Supongamos que en el plano rayado con líneas paralelas, que están una de otra a la distancia L, se lanza al azar una «aguja», o sea un segmento de longitud l, $l \leqslant L$. ¿Cuál será la probabilidad de que el segmento lanzado corte a una de las líneas existentes?

Designemos por ξ_1 el ángulo de inclinación del segmento «caído» respecto a la dirección de las líneas, por ξ_2 la distancia de su extremo *inferior hasta la línea más próxima* de arriba (fig. 6,*a*). Las condiciones de la prueba son tales que la magnitud aleatoria ξ_1 está *distribuida uniformemente*, en el segmento $[0, \pi]$, lo mismo que la magnitud aleatoria ξ_2 en el segmento $[0, L]$. Intuitivamente está claro que las magnitudes ξ_1 y ξ_2 no están ligadas entre sí; consideramos que el punto (ξ_1, ξ_2) cae en tal o cual zona A en el plano con la misma probabilidad, como si este punto (ξ_1, ξ_2) se lanzase al azar sobre el rectángulo $\Omega = \{0 \leqslant x_1 \leqslant \pi, \ 0 \leqslant x_2 \leqslant L\}$. Evidentemente, el segmento de longitud l corta una de las líneas cuando, y sólo cuando

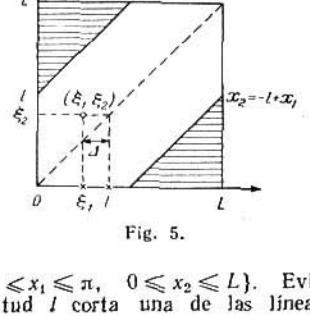

Fig. 5.

$$\xi_2 \leqslant l \operatorname{sen} \xi_1,$$

es decir, cuando el punto correspondiente (ξ_1, ξ_2) caiga en la zona A limitada por arriba con la curva $x_2 = l \operatorname{sen} x_1$ (fig. 6, *b*). De acuerdo

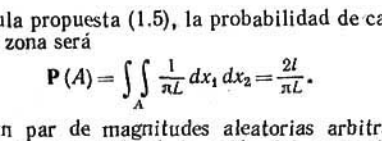

Fig. 6.

a nuestra fórmula propuesta (1.5), la probabilidad de caída del punto (ξ_1, ξ_2) en esta zona será

$$\mathbf{P}(A) = \iint\limits_A \frac{1}{\pi L}\, dx_1\, dx_2 = \frac{2l}{\pi L}.$$

Tomemos un par de magnitudes aleatorias *arbitrarias* ξ_1 y ξ_2 para las cuales la probabilidad de caída del punto (ξ_1, ξ_2) en tal o cual zona A está determinada por la igualdad (1.5), donde $p_{\xi_1\xi_2}(x_1, x_2)$, $-\infty < x_1, x_2 < \infty$, es una función no negativa, que satisface a la condición $\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} p_{\xi_1\xi_2}(x_1, x_2)\, dx_1\, dx_2 = 1$. A esta

función la llamaremos *densidad de probabilidad* de las magnitudes aleatorias $\xi_1 \xi_2$.

Supongamos, que nos interesan determinadas magnitudes $\eta_1 = \varphi_1(\xi_1, \xi_2)$ y $\eta_2 = \varphi_2(\xi_1, \xi_2)$, que son funciones de las magnitudes aleatorias iniciales ξ_1, ξ_2. Sea

$$y_1 = \varphi_1(x_1, x_2), \quad y_2 = \varphi_2(x_1, x_2)$$

la transformación biunívoca del plano con un jacobiano distinto de cero

$$|J| = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \frac{\partial \psi_1}{\partial x_2} \\ \frac{\partial \varphi_2}{\partial x_1} & \frac{\partial \varphi_2}{\partial x_2} \end{vmatrix}$$

y

$$x_1 = \psi_1(y_1, y_2), \quad x_2 = \psi_2(y_1, y_2)$$

la transformación inversa.

En la transformación examinada, la zona con el límite liso a trozos se transforma en una zona del mismo tipo. Para cualquier zona B, la caída del punto (η_1, η_2) en ella significa la caída del punto correspondiente (ξ_1, ξ_2) en la zona A, que es la imágen recíproca de B en la representación $y_1 = \varphi_1(x_1, x_2)$, $y_2 = \varphi_2(x_1, x_2)$, es decir, los sucesos $\{(\eta_1, \eta_2) \in B\}$ y $\{(\xi_1, \xi_2) \in A\}$ coinciden. Por consiguiente, la probabilidad $\mathbf{P}(B)$ de caída del punto (η_1, η_2) en la zona B, de acuerdo con la fórmula (1.5) será

$$\mathbf{P}(B) = \iint_A p_{\xi_1 \xi_2}(x_1, x_2)\, dx_1\, dx_2,$$

de donde después del cambio de variables $x_1 = \psi_1(y_1, y_2)$, $x_2 = \psi_2(y_1, y_2)$ obtenemos la fórmula del mismo tipo que (1.5):

$$\mathbf{P}(B) = \iint_B p_{\eta_1 \eta_2}(y_1, y_2)\, dy_1\, dy_2,$$

donde

$$p_{\eta_1 \eta_2}(y_1, y_2) = p_{\xi_1 \xi_2}[\psi_1(y_1, y_2), \psi_2(y_1, y_2)]\, |J|^{-1} \qquad (1.7)$$

es, según la definición dada anteriormente, *la densidad de probabilidad* de las magnitudes aleatorias η_1, η_2:

$$p_{\eta_1 \eta_2}(y_1, y_2) \geqslant 0, \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\eta_1 \eta_2}(y_1, y_2)\, dy_1\, dy_2 = 1.$$

2. Espacio de los sucesos elementales. Al examinar un par de magnitudes aleatorias ξ_1, ξ_2 con una densidad de probabilidad $p_{\xi_1 \xi_2}(x_1, x_2)$ consideramos que nos ocupamos con una prueba donde se observa

un punto aleatorio (ξ_1, ξ_2). Los «resultados elementales» de tal prueba consistente cada uno de ellos en que las magnitudes aleatorias ξ_1, ξ_2 que toman determinados valores x_1, x_2 pueden ser descritos por los puntos correspondientes (x_1, x_2) en el plano, y todos los sucesos posibles del tipo $\{(\xi_1, \xi_2) \in A\}$, o sea, «el punto aleatorio (ξ_1, ξ_2) cae en la zona A», en su esencia se describen por los conjuntos correspondientes A. Sin detenernos en los detalles físicos de las condiciones de la prueba, se puede describir el «mecanismo aleatorio» que actúa sobre ella, determinando las probabilidades $\mathbf{P}(A)$ de todos los sucesos posibles A (véase la fórmula (1.5).)

En este ejemplo claramente se manifiestan todos los principales rasgos del modelo teórico-probabilístico que es en un amplio sentido una prueba con resultados aleatorios. Al describir tal modelo se indican todos los resultados elementales posibles ω, que satisfacen a la siguiente exigencia: al final de la prueba se obtiene uno y sólo uno de estos resultados (ellos también se llaman *sucesos elementales* y todo el conjunto Ω se llama *espacio de los sucesos elementales*). Además de esto, se describe de un modo correspondiente el mecanismo de aleatoriedad en la prueba dada, es decir, se dan precisamente las probabilidades $\mathbf{P}(A)$.

$$0 \leqslant \mathbf{P}(A) \leqslant 1, \tag{1.8.}$$

de tales o cuales sucesos A; con ello la probabilidad del suceso cierto se considera igual a 1, la probabilidad del suceso imposible es igual a 0 (más adelante se hablará con más detalle de qué modo se dan las probabilidades).

Cualquier suceso A tomado por separado sólo nos interesa desde el punto de vista si ocurre como resultado de la prueba dada o no ocurre; desde este punto de vista podemos no hacer diferencia entre los sucesos A_1 y A_2, para los cuales la aparición de A_1 también significa la realización de A_2, y viceversa, la aparición de A_2 también significa la realización de A_1.

Si examinamos el suceso A[1]) ligado con la prueba Ω (como se hizo en ejemplos descritos anteriormente) se puede separar el conjunto de todos aquellos resultados elementales $\omega \in \Omega$, que conducen a la *aparición del suceso A; designemos este conjunto por el mismo símbolo A*, como en el caso del suceso examinado. Es evidente que la aparición del suceso $\{\omega \in A\}$, que consiste en que el resultado ω de nuestra prueba que entra en el conjunto separado $A \subseteq \Omega$ significa *exactamente la aparición del propio suceso A*.

En efecto, si ocurre el suceso A, entonces tiene lugar algún resultado ω que conduce a la aparición de este suceso, y todos estos resultados componen el conjunto $A \subseteq \Omega$, de modo que $\omega \in A$; por otro

[1]) Es decir, aquel suceso que aparece (o no aparece) a base del resultado aleatorio $\omega \in \Omega$.

lado, si ocurre el suceso $\{\omega \in A\}$, es decir, si tiene lugar cualquier resultado elemental $\omega \in A$, entonces según la propia definición del conjunto $A \subseteq \Omega$, esto significa que el resultado dado ω conduce a la aparición del suceso A.

Así pues, los sucesos A y $\{\omega \in A\}$ coinciden de hecho, y con ello, cualquier suceso ligado con la prueba examinada Ω puede ser descrito por el conjunto correspondiente A de resultados elementales $\omega \in \Omega$, cada uno de los cuales conduce a la aparición de este suceso. Se puede identificar convencionalmente el suceso A con el correspondiente conjunto $A \subseteq \Omega$. (Señalemos una vez más, que A, como un conjunto determinado de resultados elementales ω de la prueba Ω examinada, significa el suceso que ocurre cuando, y sólo cuando se realiza cualquier resultado elemental ω que pertenece al conjunto indicado $A \subseteq \Omega$.) *El suceso cierto* A que se realiza para cualquier resultado elemental $\omega \in \Omega$, se puede identificar con todo el espacio de los sucesos elementales Ω; es cómodo introducir también *el suceso imposible*, identificándolo con el conjunto vacío $\varnothing$.

Tal tratamiento desde el punto de vista de la teoría de conjuntos permite determinar, claramente, las distintas operaciones sobre los sucesos. Se llama precisamente *unión* o *suma de los sucesos* A_1 y A_2 al suceso $A_1 \cup A_2$, que significa que ocurre aunque sea uno de los sucesos A_1 o A_2 (el suceso $A_1 \cup A_2$, como conjunto de resultados elementales $\omega \in \Omega$, es la unión de los conjuntos correspondientes A_1 y A_2); análogamente se determina la unión $\bigcup\limits_k A_k$ de los sucesos $A_1, A_2 \ldots$ Se llama *intersección* o *producto de los sucesos* A_1 y A_2 el suceso $A_1 \cap A_2$ (designado también así: $A_1 \cdot A_2$), que consiste en que se realizan ambos sucesos A_1 y A_2 (como un conjunto en el espacio de los sucesos elementales Ω, el suceso $A_1 \cap A_2$ es la intersección de los conjuntos correspondientes A_1, $A_2 \subseteq \Omega$); análogamente se determina la intersección $\bigcap\limits_k A_k$ de muchos sucesos $A_1, A_2 \ldots$ Se llama *complementario* del suceso A el suceso $\overline{A}$, que significa que el propio suceso A no se realiza ($\overline{A}$ es el complemento del conjunto A en el espacio de los sucesos elementales Ω). Se llama *diferencia de los sucesos* A_1 y A_2 el suceso $A_1 \setminus A_2$, que significa que el suceso A_1 se realiza y el suceso A_2 no se realiza ($A_1 \setminus A_2$ es la diferencia de los conjuntos A_1, $A_2 \subseteq \Omega$). Luego, llamamos *incompatibles* o *mutuamente excluyentes* los sucesos A_1 y A_2 si ellos se excluyen mutuamente el uno al otro: cuando aparece A_1 no se realiza el suceso A_2 y viceversa, cuando aparece A_2 no se realiza el suceso A_1 (A_1 y A_2 como conjuntos en el espacio Ω no se cortan).

En el lenguaje de la teoría de conjuntos se pueden expresar otras relaciones entre los sucesos: por ejemplo, en vez de decir «el suceso A_1 lleva consigo al suceso A_2», se puede decir que «el suceso A_1 pertenece al suceso A_2», ya que uno y otro significan que al aparecer

cualquier resultado elemental $\omega \in \Omega$, que conduce al suceso A_1 ($\omega \in A_1$), también aparece el suceso A_2 (es decir, $\omega \in A_2$), y en la terminología de la teoría de conjuntos esto significa, que el conjunto A_1 pertenece al conjunto correspondiente A_2 : $A_1 \subseteq A_2$.

Algunas de las relaciones entre los distintos sucesos, indicadas anteriormente, están representadas gráficamente en la fig. 7, donde los sucesos elementales ω son los puntos del cuadrado Ω.

Como ejemplo, volvemos de nuevo a la prueba del lanzamiento de dos dados de juego. El suceso A, «cae una suma par de puntos»,

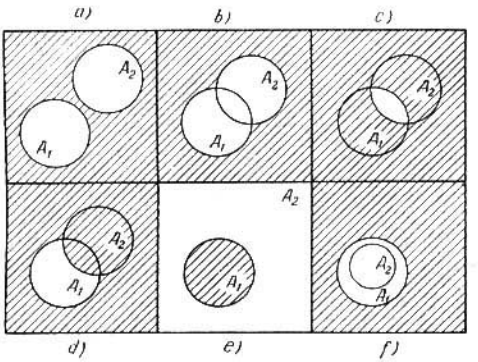

Fig. 7. *a*) A_1 y A_2 son sucesos incompatibles (que no se cortan. *b*) La figura sin rayado representa la unión $A_1 \cup A_2$. *c*) La figura sin rayado representa la intersección $A_1 \cap A_2$. *d*) La figura sin rayado representa la diferencia $A_1 \setminus A_2$. *e*) El suceso A_2 es complementario del suceso A_1. *f*) El suceso A_1 contiene al suceso A_2.

es la unión de los sucesos que no se cortan A_1, «en cada dado cae un número par de puntos» y los sucesos A_2, «en cada dado cae un número impar de puntos»; siendo $A_1 = A \setminus A_2$ y $A_2 = A \setminus A_1$. El suceso $\overline{A}$ «cae una suma impar de puntos», es complementario del suceso A. El suceso $\overline{A_1}$ «cae, aunque sea, en un dado un número impar de puntos» es complementario de A_1. El suceso $\overline{A_2}$ — «cae, aunque sea, en un dado un número par de puntos» es el suceso complementario de A_2; siendo $\overline{A_1} \setminus \overline{A} = \overline{A_1} \cap A = A_2$ y $\overline{A_2} \setminus \overline{A} = \overline{A_2} \cap A = A_1$.

No es difícil comprender que existe la siguiente ley general en los enlaces entre los distintos sucesos. Precisamente, si $A_1 \subseteq A_2$, entonces $\overline{A_2} \subseteq \overline{A_1}$; si $A = A_1 \cup A_2$, entonces $\overline{A} = \overline{A_1} \cap \overline{A_2}$; si $A = A_1 \cap A_2$, entonces $\overline{A} = \overline{A_1} \cup \overline{A_2}$. En general, si es justa una relación determinada, expresada por el signo de igualdad $=$, inclu-

sión $\subseteq$ (o $\supseteq$), unión $\cup$ e intersección $\cap$, entonces será también justa la relación, que se obtiene de la inicial sustituyendo los signos de inclusión ($\subseteq$ y $\supseteq$) por sus inversos ($\supseteq$ y $\subseteq$), la unión $\cup$ por la intersección $\cap$ y viceversa, la intersección $\cap$ por la unión $\cup$, sustituyendo al mismo tiempo los sucesos correspondientes por sus complementarios. Por ejemplo, son equivalentes las siguientes relaciones

$$\overline{A} = \overline{A}_1 \cup (A_1 \cap A_2), \quad A = \overline{\overline{A}_1 \cup (A_1 \cap A_2)},$$
$$A = A_1 \cap \overline{(\overline{A}_1 \cap A_2)}, \quad A = A_1 \cap (\overline{A}_1 \cup \overline{A}_2)$$

(observamos que aquí A coincide con la diferencia de los sucesos A_1, A_2 es decir, $A = A_1 \setminus A_2$).

3. Propiedades fundamentales de la probabilidad. Aditividad y continuidad. Examinemos cualesquiera sucesos que no se cortan A_1 y A_2, que aparecen con las probabilidades $\mathbf{P}(A_1)$ y $\mathbf{P}(A_2)$. ¿Cuál es la probabilidad del suceso $A_1 \cup A_2$, que representa su unión?

Si los sucesos examinados ligados a la prueba tienen un número finito de resultados equiprobables, entonces de acuerdo con la fórmula general (1.1) la propiedad buscada será

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2), \tag{1.9}$$

ya que el suceso $A_1 \cup A_2$ es la unión de todos los resultados elementales, que entran en A_1 o en A_2, de modo que su número es $N(A_1 \cup A_2) = N(A_1) + N(A_2)$ y

$$\frac{N(A_1 \cup A_2)}{N} = \frac{N(A_1)}{N} + \frac{N(A_2)}{N}.$$

A esa misma igualdad (1.9) llegamos, considerando a los sucesos del tipo $\{(\xi_1, \xi_2) \in A\}$, ligados con la observación de ciertas magnitudes aleatorias ξ_1 y ξ_2, cuando cada suceso de tal género se describe por el correspondiente conjunto A sobre el plano, y los sucesos que no se cortan A_1 y A_2 son en este sentido conjuntos que no se cortan. Precisamente, según la fórmula (1.5)

$$\mathbf{P}(A_1 \cup A_2) = \int\limits_{A_1 \cup A_2} \int p_{\xi_1 \xi_2}(x_1, x_2)\, dx_1\, dx_2 =$$
$$= \int\limits_{A_1} \int p_{\xi_1 \xi_2}(x_1, x_2)\, dx_1\, dx_2 + \int\limits_{A_2} \int p_{\xi_1 \xi_2}(x_1, x_2)\, dx_1\, dx_2 = \mathbf{P}(A_1) + \mathbf{P}(A_2).$$

En el modelo teórico-probabilístico la relación (1.9) para los sucesos que no se cortan A_1 y A_2, se admite en calidad de axioma; o sea, la llamada *ley de adición de las probabilidades*.

La propiedad de la aditividad expresada en ella es completamente análoga a las propiedades de longitud, superficie, volumen, etc. Debido a esta propiedad, por ejemplo, el suceso complementario $\overline{A}$

de cualquier suceso A con probabilidad $\mathbf{P}(A)$ tiene la probabilidad

$$\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$$

(la probabilidad del suceso cierto $A \cup \overline{A}$ es igual a 1);

$$\mathbf{P}(A_1) \leqslant \mathbf{P}(A_2) \quad \text{para} \quad A_1 \subseteq A_2,$$

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2),$$

para cualesquiera sucesos A_1 y A_2, etc.

Está claro, que la ley de adición de las probabilidades se extiende a cualquier número finito de sucesos que no se cortan $A_1, A_2, \ldots, A_n$; precisamente, aplicando sucesivamente la fórmula (1.9) los pares de sucesos que no se cortan A_1 y $\left(\bigcup_{k=2}^{n} A_k\right)$, A_2 y $\left(\bigcup_{k=3}^{n} A_k\right)$, $\ldots$, A_{n-1} y A_n obtenemos como resultado

$$\mathbf{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbf{P}(A_k). \tag{1.10}$$

Imaginémonos luego que se tiene una sucesión infinita de sucesos, cada uno de los cuales arrastra consigo a todos los anteriores:

$$A_1 \supseteq A_2 \supseteq \ldots \supseteq A_{n-1} \supseteq A_n \supseteq \ldots.$$

¿Qué probabilidad hay de que se realicen todos estos sucesos $A_1, A_2, \ldots$?

Según las condiciones, la realización de todos los sucesos $A_1, A_2, \ldots, A_n$ es equivalente a la aparición del suceso A_n, de modo que $\bigcap_{k=1}^{n} A_k = A_n$ y $\mathbf{P}\left(\bigcap_{k=1}^{n} A_k\right) = \mathbf{P}(A_n)$. Ya que $\mathbf{P}(A_{n+1}) \leqslant \mathbf{P}(A_n)$ cuando $A_{n+1} \subseteq A_n$, para la sucesión monótona decreciente de probabilidades $\mathbf{P}(A_n)$, $n = 1, 2, \ldots$, existe el límite $\lim_{n\to\infty} \mathbf{P}(A_n)$. En el esquema general de la teoría de probabilidades se supone que este límite coincide con la probabilidad de realización de todos estos sucesos $A_1, A_2, \ldots$, es decir, con la probabilidad del suceso $A = \bigcap_{k=1}^{\infty} A_k$:

$$\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n); \tag{1.11}$$

ésta es la propiedad llamada *continuidad de la probabilidad*.

Demostremos que la aditividad y la continuidad expresadas respectivamente por las relaciones (1.10) y (1.11) en su conjunto tienen el mismo significado que *la aditividad numerable*: para cualquier número (finito o numerable) de sucesos que no se cortan $A_1 A_2 \ldots$

$$\mathbf{P}\left(\bigcup_{k} A_k\right) = \sum_{k} \mathbf{P}(A_k). \tag{1.12}$$

En efecto, para cualesquiera de los sucesos que no se cortan A_1, A_2, ..., A_n, ..., los sucesos $B_1 = A_1$, $B_2 = A_1 \cup A_2$, ..., $B_n = \bigcup_{k=1}^{n} A_k$ forman una sucesión monótona creciente en el sentido de que $B_1 \subseteq B_2 \subseteq \ldots \subseteq B_n \subseteq \ldots$, y los sucesos complementarios $\bar{B}_1$, $\bar{B}_2$, ..., $\bar{B}_n$, ... son tales que $\bar{B}_1 \supseteq \bar{B}_2 \supseteq \ldots \supseteq \bar{B}_n \supseteq \ldots$. Con ello, el suceso $A = \bigcup_k A_k$ coincide con el suceso $\bigcup_k B_k$, y el suceso complementario $\bar{A}$ coincide con el suceso $\bigcap_k \bar{B}_k$. Según la propiedad de continuidad (1.11) $\mathbf{P}(\bar{A}) = \lim \mathbf{P}(\bar{B}_n)$. Ahora bien, $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$, $\mathbf{P}(\bar{B}_n) = 1 - \mathbf{P}(B_n)$, y según la propiedad de la aditividad (1.10) $\mathbf{P}(B_n) = \sum_{k=1}^{n} \mathbf{P}(A_k)$, $n = 1, 2, \ldots$. Por consiguiente,

$$\mathbf{P}(A) = 1 - \mathbf{P}(\bar{A}) = \lim_{n\to\infty} [1 - \mathbf{P}(\bar{B}_n)] =$$

$$= \lim_{n\to\infty} \mathbf{P}(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{P}(A_k) = \sum_{k=1}^{\infty} \mathbf{P}(A_k),$$

es decir, tiene lugar la igualdad (1.12).

Por otro lado, si $A_1 \supseteq A_2 \supseteq \ldots$ es cualquier sucesión de sucesos monótona decreciente (la aparición de A_n arrastra consigo la aparición de A_1, ..., A_{n-1}), entonces los sucesos complementarios $\bar{A}_1$, $\bar{A}_2$, ... forman una sucesión monótona creciente $\bar{A}_1 \subseteq \bar{A}_2 \subseteq \ldots$ y su unión $\bar{A} = \bigcup_k \bar{A}_k$ puede ser representada como la unión de los sucesos que no se cortan $B_1 = \bar{A}_1$, $B_2 = \bar{A}_2 \setminus \bar{A}_1$, ..., $B_n = \bar{A}_n \setminus \bar{A}_{n-1} \ldots$: $\bar{A} = \bigcup_k B_k$. De acuerdo con la fórmula (1.12)

$$\mathbf{P}(\bar{A}) = \sum_{k=1}^{\infty} \mathbf{P}(B_k) = \sum_{k=1}^{\infty} [\mathbf{P}(\bar{A}_k) - \mathbf{P}(\bar{A}_{k-1})] = \lim_{n\to\infty} \mathbf{P}(\bar{A}_n),$$

y por consiguiente, tiene lugar la relación (1.11):

$$\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Ejemplo (*juego hasta la primera pérdida*). Imaginémonos que se juega a «cara» o «cruz», cuando a cada paso el jugador gana o pierde en dependencia de si acertó o no acertó el resultado del lanzamiento de la moneda simétrica (tenemos dos resultados equiprobables «cara» o «cruz»). Supongamos que el juego se prolonga solamente hasta la primera pérdida. Cada uno de los resultados durante lanzamiento repetido n veces se puede describir por la sucesión del tipo $CCX \ldots \ldots XCX$, donde C ó X significan la caída de «cara» o «cruz» durante

el paso correspondiente; consideramos que todos los resultados posibles son equiprobables. Su número total es $N = 2^n$, de modo que la probabilidad de cada uno de los resultados, durante el lanzamiento de la moneda repetido n veces, será igual a 2^{-n}.

Para cualquier estrategia del jugador (digamos, el siempre predice «cara») el juego hasta la primera pérdida (hasta que caiga por primera vez «cruz») tiene un número infinito de resultados: la pérdida puede ser ya en el primer paso, puede ocurrir en el segundo paso y así sucesivamente. Cada resultado elemental se determina por el número n de pasos hasta la primera pérdida: $n = 1, 2, \ldots$ siendo evidentemente igual a 2^{-n} la probabilidad de que el número indicado sea precisamente n. De tal forma como modelo teórico-probabilístico del juego descrito se puede tomar el espacio de los sucesos elementales, que coincide con el conjunto de todos los números naturales $n = 1, 2, \ldots$, adjudicándole al resultado elemental n la probabilidad 2^{-n}. Como en cualquier otra prueba, la probabilidad del suceso cierto (que une a todos los resultados elementales) deberá ser igual a 1, lo que obtenemos precisamente de acuerdo con la fórmula general (1.12):

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2}} = 1.$$

Para tener un cuadro más completo se podría admitir también el resultado $n = \infty$, que significa que en la sucesión infinita de lanzamientos no cae ni una vez «cruz»; en el cuadro de nuestro esquema teórico-probabilístico tal resultado tiene una probabilidad igual a cero.

Hallemos la probabilidad de que el número de lanzamientos hasta el primer acierto con «cruz» sea par. Evidentemente, de acuerdo con la fórmula (1.12) esta probabilidad es

$$\sum_{k=1}^{\infty} \frac{1}{2^{2k}} = \frac{1}{4} \cdot \frac{1}{1 - \frac{1}{4}} = \frac{1}{3},$$

(y no 1/2 como se podría pensar equivocadamente, partiendo de la simetría de «cara» y «cruz»; a propósito, la probabilidad de que caiga «cruz» en el primer lanzamiento ya es igual a 1/2, pero ella puede aparecer por primera vez también en el lanzamiento tercero, quinto, etc).

Así pues, el modelo teórico-probabilístico representa en sí el espacio Ω de los resultados elementales ω descritos de una u otra forma, en el cual están determinadas las probabilidades $\mathbf{P}(A)$ de los sucesos examinados $A \subseteq \Omega$, que satisfacen la condición de la aditividad numerable (1.12).

El problema fundamental, que de una u otra forma surge en la mayoría de las tareas teórico-probabilísticas, consiste en lo siguiente:

se dan las probabilidades $\mathbf{P}(A)$, hablando convencionalmente, de algunos sucesos simples A, cuyo conjunto designamos por γ; se exige hallar las probabilidades $\mathbf{P}(B)$ de otros sucesos B, ligados de algún modo con los sucesos $A \in \gamma$ (más exactamente, obtenidos de los sucesos iniciales $A \in \gamma$ por medio de operaciones sucesivas de unión, intersección, paso a los sucesos complementarios, etc.).

Naturalmente, en cada caso concreto esta cuestión se resuelve distintamente, pero la pregunta general es: ¿cuándo pueden ser determinadas, en principio, las probabilidades $\mathbf{P}(B)$ a partir de las probabilidades dadas $\mathbf{P}(A)$, $A \in \gamma$. Esta cuestión importante para fundamentar la teoría de las probabilidades se resuelve positivamente, si el sistema inicial de sucesos γ posee la propiedad de que junto con los sucesos A_1, A_2 contiene también su intersección $A_1 \cap A_2$, y para cualesquiera sucesos A, $A_1 \in \gamma$ tales que A_1 pertenezca a A, la diferencia $A \setminus A_1$ se puede «desarrollar», en una suma finita de los sucesos, que no se cortan $A_2, \ldots, A_n$ de γ (tal sistema γ algunas veces se llama *desarrollable*). En el caso del sistema desarrollable γ tiene lugar la fórmula siguiente[1]):

$$\mathbf{P}(B) = \inf_{A \supseteq B} \sum_k \mathbf{P}(A_k), \tag{1.13}$$

donde el límite inferior se toma para todos los sucesos A_k de γ, $k = 1, 2, \ldots$, que dan en suma el suceso $A = \bigcup_k A_k$ que contiene a B. De ejemplo del sistema desarrollable de sucesos, en la prueba sobre la observación de una determinada magnitud aleatoria ξ, puede servir el conjunto de sucesos del tipo $A = \{x' < \xi \leqslant x''\}$, o sea, «el punto ξ cae en el segmento abierto por la izquierda $(x', x'']$»; en el ensayo sobre la observación de un par de magnitudes aleatorias ξ_1, ξ_2, se considera sistema desarrollable el conjunto de todos los sucesos del tipo $A = \{x_1' < \xi_1 \leqslant x'', x_2' < \xi_2 \leqslant x_2''\}$, o sea, «el punto con coordenadas (ξ_1, ξ_2) cae en el rectángulo $(x_1', x_1''](x_2', x_2'']$ con los lados que forman el ángulo inferior izquierdo excluidos.

En todo lo sucesivo, al examinar tal o cual modelo teórico probabilístico concreto supondremos que las probabilidades de los sucesos que nos interesan, en principio, están determinadas y la tarea consiste en hallarlas en una forma explícita.

4. Concordancia entre el modelo y la prueba. En las aplicaciones, al utilizar tal o cual modelo teórico probabilístico surge la pregunta, de cómo corresponde este modelo a la situación real de las cosas, a la prueba real.

Ejemplo (*probabilidad del nacimiento de un niño*). Se puede pensar que el nacimiento de un niño o una niña en cada uno de los casos es

[1]) Véase, por ejemplo, el libro de I. I. Guijman, A. V. Skorojod «Introducción a la teoría de los procesos aleatorios» Moscú «Nauka» 1965 (pág. 121 y en adelante).

un suceso equiprobable (es decir, el nacimiento de un niño, desde el punto de vista de la probabilidad, es análogo a la prueba del lanzamiento de la moneda: con la probabilidad 1/2 nace un niño y con la misma probabilidad nace una niña).

Examinaremos cada nacimiento como una prueba, durante la cual nace un niño con la probabilidad p y nace una niña con la probabilidad $1 - p$, con ello consideraremos que el resultado de una prueba de ningún modo influye en el resultado de la otra (en otras palabras, examinaremos los nacimientos como pruebas independientes).

Durante estas n «pruebas» el número m de niños nacidos es aleatorio. La probabilidad de que *la frecuencia* m/n del nacimiento de un niño se desvíe de $p = 1/2$ en un valor $\delta > 0$, es igual a (véase más adelante el § 3, cap. II)

$$\mathbf{P}\left\{\frac{m}{n} - p > \delta\right\} \approx 1 - \Phi(2\delta\sqrt{n}),$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\,du$, además, en la igualdad aproximada indicada el error no supera a $1/\sqrt{n}$. Elegimos δ de tal modo, que $1 - \Phi(2\delta\sqrt{n}) \leqslant \varepsilon$, donde ε es un número tan pequeño que se debe despreciar la posibilidad de aparición del suceso que tenga una probabilidad no mayor que ε. (Por ejemplo, para $\varepsilon = 0{,}002$, se puede tomar $x = 2\delta\sqrt{n} \geqslant 3$ de las tablas de la función $\Phi(x)$; véase la pág. 101.)

En casi todos los países ya hace mucho tiempo que se registra el nacimiento de cada niño de modo que ya se tiene una gran cantidad de datos estadísticos. Hablando de Suiza, desde el año 1871 hasta el 1900 nacieron 1 359 671 niños y 1 285 086 niñas[1]). De acuerdo con estos datos, para $n = 2\,644\,757$ se tuvo $m = 1\,359\,671$ y $m/n = 0{,}5141$. ¿Está esto de acuerdo con el modelo elegido, en el cual $p = 1/2$, o en realidad la probabilidad del nacimiento de un niño es mayor que 1/2?

Para $p = 1/2$ tenemos

$$\frac{m}{n} - \frac{1}{2} = 0{,}0141 > \delta = \frac{3}{2\sqrt{2\,644\,757}},$$

y de este modo, dentro del marco del modelo aceptado por nosotros (con probabilidad del nacimiento de un niño $p = 1/2$) «observamos» el suceso prácticamente imposible $\left\{\frac{m}{n} - \frac{1}{2} > \delta\right\}$, más exactamente, la probabilidad de este proceso no supera a 0,002. Al mismo tiempo,

[1]) Véase el libro de B. L. Van der Warden «Matematicheskaya Statistika» IL-Moscú 1960 (pág. 43).

si la probabilidad del nacimiento de un niño es en realidad mayor que 1/2 (digamos, $p = 0{,}51$), entonces en este suceso $\left\{\frac{m}{n} - \frac{1}{2} > \delta\right\}$ no hay nada extraño, como podemos calcular, él tiene una probabilidad próxima a 1. Basándose en estos datos, se debe d e s e c h a r la hipótesis de que $p = 1/2$ y considerar $p > \frac{1}{2}$ (en cierto sentido la mejor valuación de la probabilidad del nacimiento de un niño, según los datos estadísticos expuestos, es la f r e c u e n c i a $\frac{m}{n} = 0{,}5141$).

En nuestro curso estudiaremos los distintos modelos teórico-probabilísticos, sin examinar la cuestión en qué grado corresponden a tal o cual fenómeno real (la formulación matemática exacta y la solución de semejantes cuestiones son objeto de la ciencia especial llamada estadística matemática).

§ 2. INDEPENDENCIA Y PROBABILIDAD CONDICIONAL

1. Concepto de independencia. Examinemos dos pruebas independientes Ω_1 y Ω_2 cada una de las cuales tiene un número finito de resultados equiprobables. La independencia aquí se comprende en el sentido de que ninguno de los resultados elementales ω_1 de la primera prueba influye de ningún modo en el resultado ω_2 de la segunda prueba (y viceversa). De ejemplo puede servir (véase el comienzo del § 1) los dos lanzamientos de la moneda (ω_1 es el primer resultado, ω_2 es el segundo resultado de los lanzamientos); el lanzamiento Ω_1 de la moneda y el lanzamiento Ω_2 del dado de juego, etc.

Examinemos el suceso A_1 ligado con la primera prueba Ω_1 y el suceso A_2 ligado con la segunda prueba Ω_2. ¿Cuál será la probabilidad del suceso $A = A_1 \cdot A_2$ ligado con la prueba, que es un conjunto de pruebas independientes Ω_1, Ω_2?

El resultado elemental de las pruebas examinadas se describe por el par (ω_1, ω_2), donde ω_1 es el resultado elemental en la primera prueba, ω_2 es el resultado en la segunda prueba. La intuición indica que para los resultados independientes equiprobables $\omega_1 \in \Omega_1$ y $\omega_2 \in \Omega_2$, los resultados elementales $\omega = (\omega_1, \omega_2)$ también deben ser equiprobables. Si el número total de resultados en la primera prueba es N_1 y en la segunda prueba N_2, entonces el número total de pares posibles $\omega = (\omega_1, \omega_2)$ será evidentemente igual al producto $N_1 \cdot N_2$. De tal modo, el número de resultados elementales $\omega = (\omega_1, \omega_2)$ es $N = N_1 \cdot N_2$. El suceso $A = A_1 \cdot A_2$ significa la realización de ambos sucesos A_1 y A_2 que será, cuando, y sólo cuando tengan lugar los

resultados elementales $\omega_1 \in A_1$ y $\omega_2 \in A_2$. Evidentemente, en total se tienen $N(A) = N_1(A_1) N_2(A_2)$ resultados elementales $\omega = (\omega_1, \omega_2)$ que conducen a la aparición del suceso $A = A_1 \cdot A_2$ (aquí $N_1(A_1)$ es el número de sucesos elementales ω_1 que conducen al suceso A_1 en la primera prueba, $N_2(A_2)$ es el número de resultados elementales ω_2 que conducen al suceso A_2 en la segunda prueba. Según la fórmula general (1.1), la probabilidad del suceso $A = A_1 \cdot A_2$ se debe determinar como

$$\mathbf{P}(A) = \frac{N(A)}{N} = \frac{N_1(A_1) \cdot N_2(A_2)}{N_1 \cdot N_2},$$

que se expresa directamente a través de las probabilidades $\mathbf{P}(A_1)$, $\mathbf{P}(A_2)$ de los sucesos separados A_1, A_2 como

$$\mathbf{P}(A) = \mathbf{P}(A_1)\,\mathbf{P}(A_2). \tag{2.1}$$

Así pues, al examinar los sucesos del tipo $A = A_1 \cdot A_2$, donde los sucesos separados A_1 y A_2 están ligados con las pruebas independientes, para la probabilidad $\mathbf{P}(A)$ tendremos la fórmula (2.1).

Cierto es, que hasta ahora, sólo hemos examinado las pruebas independientes con un número finito de resultados equiprobables.

Examinemos ahora las pruebas independientes Ω_1 y Ω_2 en las que se observa la magnitud aleatoria ξ_1 en Ω_1 y la magnitud aleatoria ξ_2 en Ω_2 (las magnitudes ξ_1 y ξ_2 no están ligadas entre sí de ningún modo).

Sea $p_1(x)$ la densidad de la distribución de probabilidades de la magnitud aleatoria ξ_1 de modo que la probabilidad de cualquier suceso del tipo $A_1 = \{x_1' \leqslant \xi_1 \leqslant x_1''\}$, o sea, «el punto aleatorio ξ_1 cae en el intervalo $[x_1', x_1'']$», se da por la fórmula (1.2):

$$\mathbf{P}(A_1) = \int_{A_1} p_1(x)\,dx.$$

Representémonos mentalmente, que el segmento $[a, b]$ de los valores posibles de la magnitud aleatoria ξ_1 (este segmento puede ser también infinito) está dividido en N_1 partes de tal forma que los aciertos en cualquiera de los intervalos de división $\Delta_k = (x_k, x_{k+1}]$, $k = 1, 2, \ldots, N_1$ son sucesos equiprobables:

$$\mathbf{P}\{x_k < \xi \leqslant x_{k+1}\} = \int_{\Delta_k} p_1(x)\,dx = \frac{1}{N_1}.$$

Eligiendo a N_1 lo suficientemente grande, se puede determinar con cualquier exactitud la probabilidad del suceso A_1 (probabilidad de caída del punto aleatorio ξ_1 en el intervalo $[x_1', x_1'']$, por la fórmula siguiente:

$$\mathbf{P}(A_1) \approx \sum \int_{\Delta_k} p_1(x)\,dx \frac{N_1(A_1)}{N_1},$$

donde la suma se realiza en aquellos k, para los cuales $\Delta_k \subseteq [x_1', x_1'']$ y $N_1(A_1)$ es el número de intervalos de la división que caen en el segmento considerado $[x_1', x_1'']$. Esto significa de hecho, que podemos determinar con cualquier grado de exactitud, las probabilidades de diferentes sucesos ligados con la prueba Ω_1, examinándola convencionalmente como prueba con un número finito de resultados equiprobables, cada uno de los cuales representa en sí el acierto del punto aleatorio ξ_1 en el intervalo correspondiente de la división Δ_k, $k = 1, \ldots, N_1$.

Lo mismo se puede decir sobre la prueba Ω_2 donde se observa la magnitud aleatoria ξ_2 con la densidad de distribución de probabilidades $p_2(x)$, y se examina el suceso $A_2 = \{x_2' \leqslant \xi_2 \leqslant x_2''\}$. Utilizando la fórmula (2.1) para los sucesos con un número finito de resultados equiprobables, para valores suficientemente grandes de N_1 y N_2 (que significan el número de intervalos de la división en cada prueba por separado), tendremos con cualquier grado de exactitud: para el suceso $A = A_1 \cdot A_2$

$$\mathbf{P}(A) \approx \frac{N_1(A_1)}{N_1} \frac{N_2(A_2)}{N} \approx \mathbf{P}(A_1) \cdot \mathbf{P}(A_2).$$

Esto indica que en realidad debe tener lugar la igualdad exacta

$$\mathbf{P}(A) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2),$$

es decir, de nuevo llegamos a la fórmula (2.1), obtenida anteriormente, sólo para las pruebas independientes con un número finito de resultados equiprobables.

Examinemos, por último, los sucesos arbitrarios A_1 y A_2 ligados con algunas pruebas independientes Ω_1 y Ω_2. Igual que anteriormente, al hablar sobre la independencia, aquí tenemos en cuenta las condiciones físicas concretas, para las cuales ninguno de los resultados de la primera prueba influye de ningún modo sobre el resultado de la segunda (y viceversa).

Está claro que la probabilidad del suceso $A = A_1 \cdot A_2$ en las pruebas independientes dadas Ω_1, Ω_2 deberá ser la misma que en las pruebas independientes cualesquiera Ω_1', Ω_2', cuando nosotros examinamos cualquier suceso $A' = A_1' \cdot A_2'$, donde los sucesos $A_1' \subseteq \Omega_1'$ y $A_2' \subseteq \Omega_2'$ tienen la misma probabilidad que los sucesos tomados al principio $A_1 \subseteq \Omega_1$ y $A_2 \subseteq \Omega_2$. Ahora bien, para cualesquiera probabilidades $p_1 = \mathbf{P}(A_1)$ y $p_2 = \mathbf{P}(A_2)$ se pueden realizar las pruebas independientes Ω_1' y Ω_2' del tipo descrito anteriormente (en los que se observan determinadas magnitudes aleatorias ξ_1 y ξ_2 con densidad de distribución $p_1(x)$ y $p_2(x)$), eligiendo los sucesos correspondientes A_1' y A_2' con las mismas probabilidades $p_1 = \mathbf{P}(A_1')$ y $p_2 = \mathbf{P}(A_2')$. Anteriormente ya se mostró que debe cumplirse la relación $\mathbf{P}(A_1' \cdot A_2') = p_1 \cdot p_2$ y por consiguiente $\mathbf{P}(A_1 \cdot A_2) = p_1 \cdot p_2$.

De tal modo, la fórmula (2.1) debe tener lugar para los sucesos arbitrarios A_1, A_2, ligados con las pruebas independientes Ω_1, Ω_2.

En la teoría de probabilidades, como en cualquier otra disciplina matemática, se opera con conceptos que se formulan dentro del cuadro del modelo teórico correspondiente. El *modelo teórico probabilístico* general se describe como un determinado espacio Ω de resultados elementales ω, donde para determinar la clase de conjuntos $A \subseteq \Omega$, llamados sucesos, las probabilidades indicadas $\mathbf{P}(A)$ se someten a la regla de adición (1.12). En tal descripción no se dice una palabra sobre las determinadas «condiciones físicas concretas» cuyo análisis permitiese juzgar sobre la independencia de tales o cuales sucesos A_1, $A_2 \subseteq \Omega$.

Dentro del cuadro del modelo teórico-probabilístico general se considera señal de *independencia de los sucesos* A_1, A_2 *al cumplimiento* de la propia relación (2.1). *Se llaman independientes* precisamente, los sucesos A_1 y A_2 si

$$\mathbf{P}(A_1 \cdot A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2). \tag{2.2}$$

Ejemplo. Supongamos que de la baraja de naipes se saca al azar una carta. Tenemos una prueba concreta, que cabe en el esquema de un número N finito de resultados equiprobables (N es el número de naipes en la baraja, digamos $N = 52$). Examinemos los sucesos A_1, o sea, «se saca un naipe del palo de bastos» y A_2 «se saca un caballo». A la primera prueba le favorecen $N(A_1) = 13$ resultados elementales, ya que en el montón se tienen 13 naipes del palo de bastos, y por eso $\mathbf{P}(A_1) = 13/52 = 1/4$. Al segundo suceso le favorecen $N(A_2) = 4$ resultados elementales, ya que se tienen 4 caballos y por eso $\mathbf{P}(A_2) = 4/52 = 1/13$. Por último, al suceso $A_1 \cdot A_2$, o sea, «se saca el caballo de bastos», le favorecen, exactamente, un resultado elemental, y $\mathbf{P}(A_1 \cdot A_2) = 1/52$. Se ve que tiene lugar la relación (2.2) y, por consiguiente, los sucesos A_1 y A_2 son independientes.

Este ejemplo sencillo ya demuestra que el concepto de independencia, dado anteriormente, permite con exactitud matemática utilizarlo allí, donde, a pesar de toda la sencillez «de las condiciones físicas concretas» de la prueba examinada, es difícil abordar de otra forma la cuestión sobre la independencia de tales o cuales sucesos A_1 y A_2.

A propósito, si en el ejemplo examinado anteriormente cambiamos las condiciones de la prueba, añadiendo a los 52 naipes de la baraja corriente M naipes en blanco («carte blanche») entonces los sucesos A_1, «que sean bastos» y A_2, «que sea caballo» serán de forma evidente dependientes; para un gran M (digamos, $M = 1000$) casi con seguridad el naipe sacado será «carte blanche», y la probabilidad de sacar los bastos es muy pequeña; para la condición de que el naipe sacado sea el caballo (suceso A_2), el suceso A_1 (que el naipe sacado sean bastos)

se hace completamente real, se debe considerar que la probabilidad del suceso A_1 en las condiciones A_2 será igual a 1/4.

En general, los sucesos $A_1, A_2, \ldots$ se llaman *independientes entre sí*, (abreviadamente: *independientes*) si la probabilidad de la intersección $A_{i_1} \ldots A_{i_n}$ para cualesquiera valores distintos $i_1, \ldots, i_n$ es

$$\mathbf{P}(A_{i_1} \ldots A_{i_n}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_n}). \qquad (2.3)$$

Supongamos que se examinan determinadas pruebas Ω_1, Ω_2, con los resultados posibles $\omega_1, \omega_2, \ldots$. Se puede determinar la «prueba compleja» Ω como el espacio de los resultados elementales $\omega = (\omega_1, \omega_2, \ldots)$, donde ω_1 es el resultado elemental en Ω_1, ω_2 es el resultado elemental en Ω_2, etc. El mecanismo de aleatoriedad de tal prueba se describe corrientemente con las probabilidades $\mathbf{P}(A_{i_1} \ldots A_{i_n})$ de todas las intersecciones posibles de los sucesos $A_{i_1} \subseteq \Omega_{i_1}$, $A_{i_2} \subseteq \Omega_{i_2}, \ldots$. Si para cualesquiera sucesos $A_{i_1}, \ldots, A_{i_n}$ (ligados con diferentes pruebas $\Omega_{i_1}, \ldots, \Omega_{i_n}$), tiene lugar la relación (2.3), entonces los sucesos $\Omega_1, \Omega_2, \ldots$ se llaman *independientes*.

2. **Probabilidad condicional.** Junto al concepto de independencia, uno de los más importantes conceptos en la teoría de probabilidades es la llamada *probabilidad condicional*, que da la característica general del enlace de distintos sucesos.

Sean los sucesos A_1, A_2 ligados con la prueba Ω, que tienen un número finito N de resultados equiprobables ω. Supongamos que nos es conocida la aparición del suceso A_2 (pero no sabemos el resultado elemental ω de la prueba dada). ¿Cuál es la probabilidad del suceso A_1 en estas nuevas condiciones, cuando es conocida la aparición del suceso A_2?

Está claro, que en las nuevas condiciones pueden haber, en efecto, sólo $N(A_2)$ resultados elementales distintos (cada uno de los cuales puede conducir a la aparición del suceso A_2). Si designamos por $N(A_1 \cdot A_2)$ el número de aquellos resultados elementales del número de los resultados $\omega \in A_2$ que conducen al suceso A_1, entonces sería natural determinar la probabilidad del suceso A_1, (en condiciones de aparición del suceso A_2) como la relación $\frac{N(A_1 \cdot A_2)}{N(A_2)}$ (compárese con la fórmula general (1.1)). Pero observemos que el número $N(A_1 \cdot A_2)$ coincide con el número de todos los resultados elementales $\omega \in \Omega$ que conducen a la aparición de ambos sucesos A_1, A_2 (es decir, al suceso $A_1 \cap A_2$). Tomando en consideración, que

$$\mathbf{P}(A_1 \cdot A_2) = \frac{N(A_1 \cdot A_2)}{N}, \quad \mathbf{P}(A_2) = \frac{N(A_2)}{N},$$

podemos expresar la probabilidad del suceso A_1, determinada anteriormente, en condiciones de la aparición de A_2 (designada en adelante

por $\mathbf{P}(A_1 \mid A_2))$ por la fórmula

$$\mathbf{P}(A_1 \mid A_2) = \frac{\mathbf{P}(A_1 \cdot A_2)}{\mathbf{P}(A_2)}. \tag{2.4}$$

Esta *fórmula* también tiene sentido en el caso general (para $\mathbf{P}(A_2) > 0$); ella *determina la probabilidad condicional* $\mathbf{P}(A_1 \mid A_2)$ *del suceso* A_1 *en condiciones de la aparición del suceso* A_2.

En vista de la importancia de este concepto lo aclaramos, aún más, en un modelo. Imaginémonos que la prueba examinada consiste en la observación de un punto aleatorio ξ en un determinado espacio fásico X, actuando precisamente, el «mecanismo de aleatoriedad» del siguiente modo: la probabilidad de caída del punto aleatorio en la zona $A \subseteq X$ (suceso A) es proporcional al volumen de esta zona, que se determina como la integral por A de una función determinada positiva integrable $p(x)$:

$$\mathbf{P}(A) = \frac{1}{V} \int_A p(x)\, dx,$$

donde V es el volumen de todo el espacio fásico:

$$V = \int_X p(x)\, dx.$$

Supongamos que al observador le es conocido que el punto aleatorio ξ cae en una determinada zona A_2 (suceso A_2). ¿Cuál será en estas nuevas condiciones la aleatoriedad de caída en una u otra zona A_1?

Está claro, que bajo la condición de caída en la zona A_2, el punto ξ podrá caer en la zona A_1, sólo en el caso, cuando se tiene una intersección no vacía $A_1 \cap A_2$, siendo la caída en A_1 en las nuevas condiciones, equivalente a la caída en la intersección indicada $A_1 \cap A_2$, que se realiza con la probabilidad proporcional al volumen correspondiente $\int_{A_1 \cap A_2} p(x)\, dx$. Por consiguiente, en condiciones de la caída en la zona A_2, el punto aleatorio ξ cae en la zona A_1 con una probabilidad, igual a $\frac{1}{V(A_2)} \int_{A_1 \cap A_2} p(x)\, dx$, donde $V(A_2)$ es el volumen de la zona A_2, es decir $V(A_2) = \int_{A_2} p(x)\, dx$. Vemos que esta probabilidad $\mathbf{P}(A_1 \mid A_2)$ coincide con la relación $\frac{\mathbf{P}(A_1 \cdot A_2)}{\mathbf{P}(A_2)}$.

Partiendo de la probabilidad condicional $\mathbf{P}(A_1 \mid A_2)$, se puede introducir el siguiente concepto de independencia: el suceso A_1 no depende del suceso A_2, [si la probabilidad de aparición de A_1 no se cambia al aparecer el suceso A_2, más exactamente, si la probabilidad

condicional $\mathbf{P}(A_1 \mid A_2)$ coincide con la probabilidad inicial $\mathbf{P}(A_1)$ del suceso examinado A_1, es decir:

$$\mathbf{P}(A_1 \mid A_2) = \mathbf{P}(A_1).$$

Pero como vemos, en la fórmula (2.4) esta igualdad es equivalente a la igualdad (2.2) y de nuevo llegamos al concepto anterior de independencia de los sucesos A_1 y A_2.

En las aplicaciones, cuando se trata de hallar la probabilidad $\mathbf{P}(A)$ de tal o cual suceso A, ligado con una determinada prueba compleja Ω, es cómodo suponer condicionalmente, la realización, de uno u otro modo, del suceso elegido B, si la introducción de esta condición simplifica la prueba y permite determinar la probabilidad condicional $\mathbf{P}(A \mid B)$.

Supongamos que tenemos, precisamente, un tal sistema llamado *sistema completo* de sucesos $B_1, B_2, \ldots$ de modo que como resultado de la prueba examinada ocurre, u n o y s ó l o u n o, de los sucesos $B_1, B_2, \ldots$ (en otras palabras, estos sucesos no se cortan y su unión es el suceso cierto). Después de determinar las probabilidades condicionales $\mathbf{P}(A \mid B_k)$ del suceso A que nos interesa, en diferentes condiciones B_k, $k = 1, 2, \ldots$, se puede calcular la probabilidad $\mathbf{P}(A)$ del suceso A por la siguiente *fórmula de la probabilidad total*:

$$\mathbf{P}(A) = \sum_k \mathbf{P}(A \mid B_k)\,\mathbf{P}(B_k). \tag{2.5}$$

Esta fórmula se puede deducir fácilmente, representando el suceso A como la unión de los sucesos que se cortan $A_k = A \cdot B_k$, $k = 1, 2, \ldots$:

$$A = A \cdot \bigcup_k B_k = \bigcup_k (A \cdot B_k),$$

y utilizando la regla de adición de probabilidades (véase (1.12)), según la cual

$$\mathbf{P}(A) = \sum_k \mathbf{P}(A \cdot B_k).$$

Expresando los sumandos $\mathbf{P}(A \cdot B_k)$ a través de la probabilidad $\mathbf{P}(A \cdot B_k) = \mathbf{P}(A \mid B_k) \cdot \mathbf{P}(B_k)$, obtenemos la igualdad (2.5)

Para ilustrar las posibilidades que abre la utilización de la fórmula de la probabilidad total, detengámonos aunque sea en un ejemplo.

Ejemplo (tarea sobre el arruinamiento del jugador). Examinemos el juego llamado de «cara» o «cruz» cuando el jugador elige la «cara» o la «cruz». Si cae la cara de la moneda designada por el jugador, entonces gana recibiendo, digamos, 1 rublo; en caso contrario pierde lo mismo. Supongamos que el capital inicial del jugador consta de x rublos y el jugador se plantea el objetivo de aumentarlo hasta una determinada suma de $a > x$ rublos. El juego se termina cuando el jugador o bien reúne la suma a determinada anteriormente, o bien

se arruina, perdiendo todo el capital que tenía. ¿Cuál será la probabilidad de que, al fin o al cabo, el jugador se arruine, ya que no reunió la suma a de rublos deseada?

Es claro, que esta probabilidad depende del capital inicial x y de la suma final a. Designando por $p(x)$ la probabilidad de que, teniendo el jugador x rublos, a pesar de todo se arruina. Entonces la probabilidad de arruinamiento en condiciones de ganancia en el primer paso en nuestras designaciones será $p(x+1)$, ya que después de la ganancia el capital se hace igual a $x+1$. Análogamente, la probabilidad del arruinamiento en condiciones de pérdida en el primer paso es igual a $p(x-1)$, ya que después de la pérdida el capital del jugador quedará igual a $x-1$. Designemos por B_1 el suceso, consistente en que el jugador ganó en el primer paso, por B_2 el suceso consistente en que perdió y sea A el suceso que significa el arruinamiento del jugador. Las probabilidades condicionales del arruinamiento se expresan en las designaciones admitidas por nosotros así:

$$\mathbf{P}(A \mid B_1) = p(x+1), \quad \mathbf{P}(A \mid B_2) = p(x-1).$$

Los sucesos B_1 y B_2 forman un sistema completo, ya que en el primer paso el jugador o bien gana o pierde, siendo $\mathbf{P}(B_1) = \mathbf{P}(B_2) = 1/2$. La fórmula de la probabilidad total da la siguiente relación de las probabilidades buscadas $p(x)$:

$$p(x) = \frac{1}{2}[p(x+1) + p(x-1)]$$

para todos los $x = 1, \ldots, a-1$ (evidentemente se debe establecer que $p(0) = 1$ y $p(a) = 0$).

La solución de la ecuación

$$f(x+1) = 2f(x) - f(x-1), \quad x = 1, 2, \ldots,$$

con relación a la función $f(x)$, la cual para $1 \leqslant x \leqslant a-1$ satisface a la probabilidad $p(x)$, se determina sucesivamente a través de $y_0 = f(0)$, $y_1 = f(1)$ y para todos los $x = 2, 3, \ldots$, y, por consiguiente, puede existir sólo una solución con las «condiciones iniciales» dadas $y_0 = f(0)$, $y_1 = f(1)$. Como se puede comprobar fácilmente, esta solución tiene la forma

$$f(x) = y_0 - (y_1 - y_0)x,$$

y para $y_0 = 1$, $y_1 = p(1)$ obtenemos

$$p(x) = 1 + (1 - p(1))x,$$

de donde, teniendo en cuenta que $p(a) = 1 - (1 - p(1))a = 0$, hallamos $p(1) = 1 - \frac{1}{a}$ y finalmente obtenemos

$$p(x) = 1 - \frac{x}{a}, \quad x = 0, 1, \ldots, a.$$

§ 3. MAGNITUDES ALEATORIAS Y DISTRIBUCION DE LAS PROBABILIDADES. INDEPENDENCIA

1. Distribuciones discreta y continua. En el cuadro de tal o cual modelo teórico-probabilístico, que se describe con el espacio correspondiente de sucesos elementales Ω y con unas probabilidades determinadas $\mathbf{P}(A)$ de los sucesos $A \subseteq \Omega$, cuando se examina tal o cual magnitud aleatoria ξ, suponemos que su dependencia del suceso (más exactamente, del resultado elemental $\omega \in \Omega$) es tal, que están determinadas las probabilidades de todos los sucesos posibles del tipo $\{x' \leqslant \xi \leqslant x''\}$. Aquí tenemos presente la magnitud aleatoria real, o sea la función real $\xi = \xi(\omega)$ del resultado elemental $\omega \in \Omega$.

Inmediatamente se debe decir, que en las tareas teórico-probabilísticas, como regla, la dependencia explícita $\xi = \xi(\omega)$ de $\omega \in \Omega$ no juega un papel importante. En efecto, la característica importante de dependencia entre la magnitud ξ que nos interesa y el caso dado, nos la dan las probabilidades

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} \tag{3.1}$$

de todos los sucesos posibles $A = \{x' \leqslant \xi \leqslant x''\}$; en su conjunto, ellos muestran cómo está distribuida la probabilidad de caída del punto aleatorio ξ en tal o cual intervalo $[x', x'']$, en otras palabras, *se da la distribución de probabilidades de la magnitud aleatoria* ξ.

Al examinar una magnitud aleatoria dada ξ, se puede tomar la recta real E^1, en calidad de espacio de los sucesos elementales, fijando cada suceso elemental $\{\xi = x\}$ por el punto correspondiente $x \in E^1$, y determinando formalmente la dependencia entre ξ y el suceso elemental $x \in E^1$ por la fórmula $\xi(x) \equiv x$; el «mecanismo de aleatoriedad» de tal prueba se describe por la distribución de las probabilidades (3.1).

Señalemos una vez más, que la dependencia explícita del suceso elemental (digamos, del tipo $\xi(x) \equiv x$) no juega un papel importante desde el punto de vista del comportamiento probable de la magnitud aleatoria ξ; por ejemplo, a las mismas leyes de probabilidad se someten la magnitud aleatoria $\xi(x) \equiv x$, $-\infty < x < \infty$, con una densidad de distribución uniforme ($p_\xi(x) = \frac{1}{L}$ para $0 \leqslant x \leqslant L$, $p_\xi(x) = 0$ para $x < 0$, $x > L$), y el punto aleatorio sobre la circunferencia de longitud L, que determina la posición de la bolita en el juego a la ruleta (véase el § 1).

Supongamos que se tiene un número finito o numerable de valores de x, cada uno de los cuales la magnitud aleatoria ξ puede tomarlo con la correspondiente probabilidad

$$P_\xi(x) = \mathbf{P}\{\xi = x\} \quad \left(\sum_x P_\xi(x) = 1\right),$$

es decir, la magnitud aleatoria ξ toma uno de los valores indicados x con la probabilidad 1 (de hecho estos valores de x son los únicos posibles para ξ). Tal magnitud aleatoria se llama *discreta*; también se llama discreta su distribución de probabilidades.

Para cualesquiera x', x'' $(x' \leqslant x'')$

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = \sum_{x'}^{x''} P_\xi(x), \tag{3.2}$$

donde la suma se extiende a los límites señalados, al número finito o numerable de valores posibles de x para los cuales $P_\xi(x) > 0$.

La distribución de probabilidades de la magnitud aleatoria ξ se llama *continua*, si se tiene la densidad de probabilidad $p_\xi(x)$; es decir para cualesquiera x', x'' $(x' \leqslant x'')$ (véase (1.2))

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = \int_{x'}^{x''} p_\xi(x)\,dx \tag{3.3}$$

(la densidad de la probabilidad $p_\xi(x)$, o de otro modo: *la densidad de distribución de probabilidades* es tal función no negativa e integrable para la cual $\int_{-\infty}^{+\infty} p_\xi(x)\,dx = 1$).

La magnitud ξ, distribuida continuamente, toma cada uno de los valores fijados de x, sólo con una probabilidad igual a cero:

$$\mathbf{P}\{\xi = x\} = \lim_{\substack{x' \to x-0 \\ x'' \to x+0}} \int_{x'}^{x''} p_\xi(x)\,dx = 0,$$

y para cualquier punto de continuidad de la función $p_\xi(x)$ tenemos:

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} \sim p_\xi(x) \cdot \Delta x$$

para $\Delta x = x'' - x' \to 0$, $x' \leqslant x \leqslant x''$.

Naturalmente, la magnitud aleatoria ξ puede no relacionarse ni al tipo discreto, ni al continuo (véase el ejemplo de la pág. 42). Se puede dar la distribución general de las probabilidades de la magnitud aleatoria ξ con ayuda de la *llamada función de distribución*

$$F_\xi(x) = \mathbf{P}\{\xi \leqslant x\}, \quad -\infty < x < \infty \tag{3.4}$$

y, precisamente, para cualesquiera x', x'' $(x' \leqslant x'')$

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = F_\xi(x'') - F_\xi(x' - 0),$$

donde $F_\xi(x - 0)$ significa el límite $\lim_{h \to 0} F_\xi(x - h)$ para $h > 0$.

La función $F_\xi(x)$ que satisface a la igualdad (3.4) es no negativa, monótona creciente (no decreciente) y continua por la derecha

$(F_\xi(x) = F_\xi(x+0))$, siendo

$$\lim_{x\to-\infty} F_\xi(x) \quad 0, \quad \lim_{x\to\infty} F_\xi(x) \quad 1.$$

Las propiedades enumeradas se pueden deducir, fácilmente de la aditividad y continuidad de las probabilidades. Precisamente, el suceso $\{\xi \leqslant x\}$ es la intersección de los sucesos monótonos decrecientes $\{\xi \leqslant x_n\}$, donde x_n, $n = 1, 2, \ldots$, es cualquier sucesión monótona decreciente, que converge en x, y según la propiedad de continuidad de la probabilidad $\lim_{n\to\infty} \mathbf{P}\{\xi \leqslant x_n\} = \mathbf{P}\{\xi \leqslant x\}$; según esta misma propiedad, para los sucesos $\{\xi \leqslant x_n\}$, $x_n \to -\infty$, que dan en su intersección el suceso imposible (vacío) $\lim_{n\to\infty} \mathbf{P}\{\xi \leqslant x_n\} = 0$; para los sucesos monótonos crecientes $\{\xi \leqslant x_n\}$, $x_n \to +\infty$, que dan en su suma el suceso cierto $\{\xi < +\infty\}$, $\lim_{n\to\infty} \mathbf{P}\{\xi \leqslant x_n\} = 1$. Luego, el suceso $\{\xi \leqslant x''\}$ es la unión de los sucesos que no se cortan $\{\xi \leqslant x\}$ y $\{x < \xi \leqslant x''\}$, de modo que

$$\mathbf{P}\{x < \xi \leqslant x''\} = \mathbf{P}\{\xi \leqslant x''\} - \mathbf{P}\{\xi \leqslant x\}$$

y

$$\mathbf{P}(x' \leqslant \xi \leqslant x'') = \lim_{x\to x'-0} \mathbf{P}\{x < \xi \leqslant x''\} = F_\xi(x'') - F_\xi(x'-0).$$

Señalemos que para la magnitud discreta ξ

$$F_\xi(x) = \sum_{y\leqslant x} P_\xi(y),$$

y para la magnitud aleatoria distribuida continuamente

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(y)\,dy,$$

donde $p_\xi(x)$ es la d e n s i d a d d e p r o b a b i l i d a d correspondiente, que coincide en sus puntos de continuidad con la derivada de la función de distribución:

$$p_\xi(x) = F'_\xi(x). \tag{3.5}$$

2. Distribución conjunta de las probabilidades. Examinemos dos magnitudes aleatorias ξ_1 y ξ_2 dependientes del resultado elemental $\omega \in \Omega$ de una misma prueba.

En calidad de modelo formal se puede tomar la prueba, consistente en la observación de las magnitudes aleatorias ξ_1, ξ_2, y más exactamente, en calidad de espacio de los sucesos elementales se puede tomar el plano real E^2, fijando cada suceso elemental $\{\xi_1 = x_1, \xi_2 = x_2\}$ con el punto correspondiente $(x_1, x_2) \in E^2$ y determinando, formalmente, la dependencia entre las magnitudes ξ_1, ξ_2 y el resultado ele-

mental $(x_1, x_2) \in E^2$ por la fórmula $\xi_1(x_1, x_2) \equiv x_1$, $\xi_2(x_1, x_2) \equiv x_2$; el «mecanismo de aleatoriedad» de tal prueba se da por *la distribución conjunta de las probabilidades*, es decir, las probabilidades

$$\mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1''; \quad x_2' \leqslant \xi_2 \leqslant x_2''\}$$

de todos los sucesos posibles $A = \{x_1' \leqslant \xi_1 \leqslant x_1''; \ x_2' \leqslant \xi_2 \leqslant x_2''\}$.

Señalemos que ξ_1, ξ_2 se pueden considerar como coordenadas de una magnitud vectorial aleatoria $\xi = (\xi_1, \xi_2)$ con una distribución de probabilidades, dadas por las probabilidades de caída del punto ξ en las distintas zonas rectangulares del tipo $A = [x_1', x_1''] \times [x_2', x_2'']$:

$$\mathbf{P}\{\xi \in A\} = \mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1''; \ x_2' \leqslant \xi_2 \leqslant x_2''\}.$$

Para las magnitudes discretas (ξ_1, ξ_2) la distribución de probabilidades se dan por las probabilidades

$P_{\xi_1, \xi_2}(x_1, x_2) = \mathbf{P}\{\xi_1 = x_1, \ \xi_2 = x_2\}$, $-\infty < x_1, \ x_2 < \infty$,

y precisamente,

$$\mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1''; \ x_2' \leqslant \xi_2 \leqslant x_2''\} = \sum_{x_1'}^{x_1''} \sum_{x_2'}^{x_2''} P_{\xi_1, \xi_2}(x_1, x_2), \tag{3.6}$$

donde en los límites indicados la suma se realiza por el número finito o numerable de valores (x_1, x_2) para los cuales $P_{\xi_1, \xi_2}(x_1, x_2) > 0$.

Las distribuciones de las probabilidades de cada una de estas magnitudes ξ_1, ξ_2 pueden ser determinadas por separado según las fórmulas

$$\left.\begin{aligned} P_{\xi_1}(x_1) &= \sum_{-\infty < x_2 < \infty} P_{\xi_1, \xi_2}(x_1, x_2) \\ \text{y} \qquad & \\ P_{\xi_2}(x_2) &= \sum_{-\infty < x_1 < \infty} P_{\xi_1, \xi_2}(x_1, x_2), \end{aligned}\right\} \tag{3.7}$$

que se obtienen de (3.6) para $x_2' = -\infty$, $x_2'' = \infty$ y $x_1' = -\infty$, $x_1'' = \infty$.

Se dice que las magnitudes aleatorias ξ_1, ξ_2 tienen *una distribución conjunta continua de probabilidades*, si existe *la densidad de probabilidad*, o de otro modo, *la densidad de la distribución conjunta* $p_{\xi_1 \xi_2}(x_1, x_2)$ es una función no negativa integrable, de un par de variables $(x_1, x_2) \in \in E^2$,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2 = 1,$$

de modo que

$$\mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1''; \ x_2' \leqslant \xi_2 \leqslant x_2''\} = \int_{x_1'}^{x_1''} \int_{x_2'}^{x_2''} p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2. \tag{3.8}$$

La fórmula (3.8) se puede volver a escribir en la forma

$$\mathbf{P}\{(\xi_1, \xi_2) \in A\} = \iint\limits_A p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2. \tag{3.9}$$

Mostremos, que ésta se puede extender desde los rectángulos $A = [x_1', x_1''] \times [x_2', x_2'']$ a las zonas de estructura más compleja, por ejemplo, a cualesquiera zonas A con frontera lisa a trozos.

Observemos primeramente, que la probabilidad de caer en el rectángulo cerrado $[x_1', x_1''] \times [x_2', x_2'']$ y en el rectángulo semiabierto $(x_1', x_1''] \times (x_2', x_2'']$ es una misma, ya que la probabilidad de que el punto aleatorio (ξ_1, ξ_2) del plano E^2 caiga en la frontera del rectángulo es igual a cero. Esto se puede deducir, fácilmente, de la fórmula (3.8) y de la propiedad de continuidad de la probabilidad.

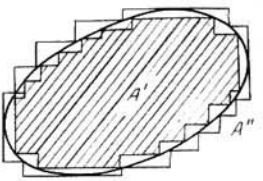

Fig. 8.

Pero tratándose del rectángulo semiabierto del tipo $(x_1', x_1''] \times (x_2', x_2'']$, utilizando la ley de adición de probabilidades, podemos extender la fórmula (3.8) a los polígonos A, que admiten la división en rectángulos que no se cortan A_k del tipo indicado:

$$\mathbf{P}\{(\xi_1, \xi_2) \in A\} = \sum_k \mathbf{P}\{(\xi_1, \xi_2) \in A_k\} =$$

$$\sum_k \int\limits_A \int\limits_{A_k} p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2 = \int\limits_A \int p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2.$$

Pongamos atención a la circunstancia de que los rectángulos semiabiertos A_k quedan s i n c o r t a r s e, incluso al tomar contacto uno con otro (lo que no se puede decir de los rectángulos cerrados).

A continuación, examinando ahora la zona A con frontera lisa a trozos, podemos inscribir en ella el polígono A' ($A' \subseteq A$) (fig. 8) del tipo examinado anteriormente, de modo que para cualquier valor dado de $\varepsilon > 0$

$$\mathbf{P}\{\xi \in A'\} = \int\limits_{A'} \int p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2 \geqslant$$

$$\geqslant \int\limits_A \int p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2 - \varepsilon.$$

Ya que para $A' \subseteq A$ el suceso $\{\xi \in A'\}$ pertenece al suceso $\{\xi \in A\}$, entonces $\mathbf{P}\{\xi \in A\} \geqslant \mathbf{P}\{\xi \in A'\}$ y tanto más

$$\mathbf{P}\{\xi \in A\} \geqslant \int\limits_A \int p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2 - \varepsilon.$$

De modo completamente análogo, utilizando el correspondiente polígono A'', circunscrito junto a la zona A ($A \subseteq A''$), se puede obtener la desigualdad

$$\mathbf{P}\{\xi \in A\} \leqslant \int_A \int p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2 + \varepsilon.$$

Debido a la arbitrariedad de ε, de aquí sacamos la conclusión de que en realidad

$$\mathbf{P}(\xi \in A) = \int_A \int p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1\, dx_2,$$

lo que se exigía demostrar.

Señalemos, que si las magnitudes aleatorias ξ_1, ξ_2 tienen una distribución conjunta de probabilidades con la densidad $p_{\xi_1, \xi_2}(x_1, x_2)$, entonces, cada una de ellas por separado también tiene densidad de distribución de probabilidades; precisamente, utilizando la fórmula (3.8), para $x_2' = -\infty$, $x_2'' = \infty$ y $x_1' = -\infty$, $x_1'' = \infty$, no es difícil comprobar que

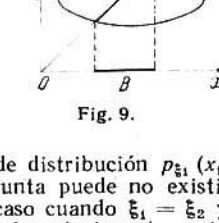

Fig. 9.

$$\left.\begin{aligned} p_{\xi_1}(x_1) &= \int_{-\infty}^{+\infty} p_{\xi_1, \xi_2}(x_1, x_2)\, dx_2 \\ \text{y} \qquad p_{\xi_2}(x_2) &= \int_{-\infty}^{+\infty} p_{\xi_1, \xi_2}(x_1, x_2)\, dx_1. \end{aligned}\right\} \quad (3.10)$$

La afirmación inversa, hablando en general es falsa. Para las magnitudes aleatorias ξ_1 y ξ_2 que tienen por separado las densidades de distribución $p_{\xi_1}(x_1)$ y $p_{\xi_2}(x_2)$, la densidad de distribución conjunta puede no existir. Por ejemplo, así será, precisamente, en el caso cuando $\xi_1 = \xi_2$ y $\mathbf{P}\{(\xi_1, \xi_2) \in A\} = 0$ para cualquier zona A sobre el plano (x_1, x_2), que no se corta con la recta $x_1 = x_2$; para la zona arbitraria A

$$\mathbf{P}\{(\xi_1, \xi_2) \in A\} = \int_B p_{\xi_1}(x)\, dx,$$

donde B significa un conjunto sobre el eje x_1, que se obtiene como imagen de la intersección A con la recta $x_1 = x_2$ en su representación $(x_1, x_2) \to x_1$ (fig. 9).

Las magnitudes aleatorias ξ_1 y ξ_2 se llaman *independientes*, si para cualesquiera intervalos $[x_1', x_1'']$ y $[x_2', x_2'']$ los sucesos $\{x_1' \leqslant \xi_1 \leqslant x_1''\}$ y $\{x_2' \leqslant \xi_2 \leqslant x_2''\}$ son independientes, es decir,

$$\mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1'';\ x_2' \leqslant \xi_2 \leqslant x_2''\} = \mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1''\} \times \mathbf{P}\{x_2' \leqslant \xi_2 \leqslant x_2''\}.$$

Como se ve fácilmente, la independencia de las magnitudes discretas ξ_1 y ξ_2 significa, que la distribución conjunta de probabilidades se determina por la fórmula

$$P_{\xi_1,\xi_2}(x_1,\ x_2) = P_{\xi_1}(x_1)\cdot P_{\xi_2}(x_2). \tag{3.11}$$

En caso de la distribución continua de las magnitudes independientes ξ_1 y ξ_2

$$\mathbf{P}\{x_1' \leqslant \xi_1 \leqslant x_1'';\ x_2' \leqslant \xi_2 \leqslant x_2''\} = \int_{x_1'}^{x_1''} p_{\xi_1}(x_1) \times$$

$$\times\, dx_1 \cdot \int_{x_2'}^{x_2''} p_{\xi_2}(x_2)\, dx_2 = \int_{x_1'}^{x_1''}\int_{x_2'}^{x_2''} [p_{\xi_1}(x_1)\, p_{\xi_2}(x_2)]\, dx_1\, dx_2$$

para todos los intervalos $[x_1', x_1'']$ y $[x_2', x_2'']$. Vemos que se tiene una densidad de distribución conjunta, y precisamente

$$p_{\xi_1,\xi_2}(x_1,\ x_2) = p_{\xi_1}(x_1)\, p_{\xi_2}(x_2). \tag{3.12}$$

También se ve fácilmente, que si la densidad de distribución conjunta $p_{\xi_1,\xi_2}(x_1,\ x_2)$ se expresa por la fórmula (3.12), entonces, las magnitudes aleatorias ξ_1 y ξ_2 son independientes.

3. Transformación de las magnitudes aleatorias. Examinemos la cuestión de cómo varía la distribución de probabilidades de las magnitudes aleatorias durante tal o cual transformación.

Anteriormente, en el § 1 se mostró que para la magnitud aleatoria ξ con densidad de probabilidad $p_\xi(x)$, durante la transformación mutuamente unívoca $y = \varphi(x)$, donde $\varphi'(x) \neq 0$, la densidad de probabilidad $p_\eta(y)$ de la magnitud $\eta = \varphi(\xi)$ se determina por la fórmula (1.4). El método de sustitución de la variable, utilizado al deducir esta fórmula, también se puede utilizar en el caso cuando en determinados puntos $\varphi'(x) = 0$ y la transformación $y = \varphi(x)$ no es mutuamente unívoca.

Ejemplo. Examinemos la transformación $y = x^2$. Evidentemente, para la magnitud aleatoria $\eta = \xi^2$ tenemos (fig. 10)

$$\mathbf{P}\{y' \leqslant \eta \leqslant y''\} = \mathbf{P}\{-x'' \leqslant \xi \leqslant -x'\} + \mathbf{P}\{x' \leqslant \xi \leqslant x''\} =$$

$$= \int_{-x''}^{-x'} p_\xi(x)\, dx + \int_{x'}^{x''} p_\xi(x)\, dx = \int_{x'}^{x''} [p_\xi(-x) + p_\xi(x)]\, dx =$$

$$= \int_{y'}^{y''} [p_\xi(-\sqrt{y}) + p_\xi(\sqrt{y})]\, \frac{1}{2\sqrt{y}}\, dy,$$

y de tal modo, la densidad de probabilidad de la magnitud aleatoria η es

$$p_\eta(y) = \begin{cases} [p_\xi(-\sqrt{y}) + p_\xi(\sqrt{y})]\frac{1}{2\sqrt{y}}, & y \geqslant 0, \\ 0, & y < 0. \end{cases}$$

En el § 1 fue mostrado que para las magnitudes aleatorias (ξ_1, ξ_2) con densidad de probabilidad $p_{\xi_1,\xi_2}(x_1, x_2)$ durante la transformación mutuamente unívoca del plano $y_1 = \varphi_1(x_1, x_2)$, $y_2 = \varphi_2(x_1, x_2)$ con un jacobiano no degenerado

$$|I| = \begin{vmatrix} \frac{\partial\varphi_1}{\partial x_1} & \frac{\partial\varphi_1}{\partial x_2} \\ \frac{\partial\varphi_2}{\partial x_1} & \frac{\partial\varphi_2}{\partial x_2} \end{vmatrix}$$

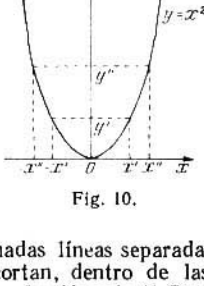

Fig. 10.

la densidad de probabilidad $p_{\eta_1,\eta_2}(y_1, y_2)$ de las magnitudes aleatorias $\eta_1 = \varphi_1(\xi_1, \xi_2)$, $\eta_2 = \varphi_2(\xi_1, \xi_2)$ se determina por la fórmula (1.7). Esta fórmula también tiene sentido para la transformación, en la cual el jacobiano $|I|$ se convierte en cero en determinadas líneas separadas, que dividen el plano E^2 en zonas que no se cortan, dentro de las cuales el jacobiano $|I|$ no se hace cero.

La fórmula (1.7) también se puede utilizar cuando se trata de la transformación del tipo $\eta = \varphi(\xi_1, \xi_2)$. Si se puede introducir, precisamente, la transformación auxiliar $\eta_1 = \eta$, $\eta_2 = \varphi_2(\xi_1, \xi_2)$ que conduce a la densidad $p_{\eta_1,\eta_2}(y_1, y_2)$ del tipo (1.7), entonces la densidad de probabilidad $p_\eta(y)$ de la magnitud η será (véase (3.10))

$$p_{\eta_1}(y_1) = \int_{-\infty}^{\infty} p_{\xi_1,\xi_2}[\psi_1(y_1, y_2), \psi_2(y_1, y_2)]\,|I|^{-1}\,dy_2. \qquad (3.13)$$

Ejemplo. Sean ξ_1 y ξ_2 magnitudes aleatorias independientes con densidades de distribución de probabilidades $p_1(x)$ y $p_2(x)$. Hallemos la distribución de probabilidades de su relación ξ_1/ξ_2.

Observemos primeramente, que ya que $\xi_2 = 0$ solamente con una probabilidad igual a cero, la magnitud $\eta_1 = \xi_1/\xi_2$ toma valores finitos con una probabilidad 1. Examinemos la transformación

$$y_1 = \frac{x_1}{x_2}, \quad y_2 = x_2 \left(\eta_1 = \frac{\xi_1}{\xi_2}, \ \eta_2 = \xi_2\right)$$

con el jacobiano $|I|^{-1} = |y_2|$. La densidad de probabilidad $p_{\eta_1,\eta_2}(y_1, y_2)$ de las magnitudes η_1, η_2 en cada una de las zonas $y_2 < 0$

y $y_2 > 0$ se puede determinar por la fórmula (1.7)

$$p_{\eta_1, \eta_2}(y_1, y_2) = \begin{cases} p_1(y_1 \cdot y_2)\, p_2(y_2)\, y_2 \text{ para } y_2 > 0 \\ -p_1(y_1 \cdot y_2)\, p_2(y_2)\, y_2 \text{ para } y_2 < 0. \end{cases}$$

Teniendo en cuenta que la densidad de probabilidad de la magnitud η_1 tomada por separado es

$$p_{\eta_1}(y_1) = \int_{-\infty}^{\infty} p_{\eta_1, \eta_2}(y_1, y_2)\, dy_2, \quad -\infty < y_1 < \infty,$$

en el caso considerado tenemos

$$p_{\eta_1}(y) = \int_0^{\infty} p_1(yx)\, p_2(x)\, x\, dx - \int_{-\infty}^{0} p_1(yx)\, p_2(x)\, x\, dx. \qquad (3.14)$$

Ejemplo. Sean ξ_1 y ξ_2 magnitudes aleatorias *independientes* con densidades de probabilidad $p_1(x)$ y $p_2(x)$. ¿Cuál será la distribución de probabilidades de su suma $\xi_1 + \xi_2$?

La transformación $\varphi_1(x_1, x_2) = x_1 + x_2$ se puede complementar hasta la transformación no degenerada del plano, suponiendo que $\varphi_2(x_1, x_2) = x_2$. Con ello $|I| = 1$ y $\psi_2(y_1, y_2) = y_2$, $\psi_1(y_1, y_2) = y_1 - y_2$. Para las magnitudes independientes ξ_1, ξ_2 la densidad conjunta de probabilidad es $p_{\xi_1, \xi_2}(x_1, x_2) = p_1(x_1)\, p_2(x_2)$ y para las magnitudes $\eta_1 = \xi_1 + \xi_2$, $\eta_2 = \xi_2$ la densidad de probabilidad será $p_{\eta_1, \eta_2}(y_1, y_2) = p_1(y_1 - y_2) \cdot p_2(y_2)$. Por consiguiente, la densidad buscada $p_{\eta_1}(y)$ es

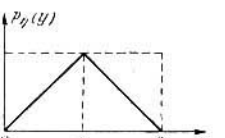

Fig. 11.

$$p_{\eta_1}(y) = \int_{-\infty}^{\infty} p_1(y - x)\, p_2(x)\, dx. \qquad (3.15)$$

Esta expresión de la densidad de distribución de probabilidades de la magnitud aleatoria $\xi_1 + \xi_2$ se llama *envoltura* o *composición* de las densidades $p_1(x)$ y $p_2(x)$.

Sean, en caso particular, ξ_1, ξ_2 magnitudes aleatorias independientes con distribución uniforme en el segmento [0, 1]. La densidad de cada una de ellas es

$$p(x) = \begin{cases} 1 \text{ para } 0 \leqslant x \leqslant 1, \\ 0 \text{ para } x < 0,\ x > 1, \end{cases}$$

y para la suma $\eta = \xi_1 + \xi_2$ obtenemos de la fórmula (3.15) el tal llamado *triángulo de distribución* con densidad (fig. 11)

$$p_\eta(y) = \begin{cases} \int\limits_0^y dy = y & \text{para } 0 \leqslant y \leqslant 1, \\ \int\limits_{y-1}^1 dx = 2 - y & \text{para } 1 \leqslant y \leqslant 2 \\ 0 & \text{para } y < 0,\ y > 2. \end{cases}$$

4. Distribuciones condicionales de probabilidades. El enlace entre distintas magnitudes ξ y η (o de la magnitud ξ con un determinado suceso B), se puede caracterizar por la tal llamada *distribución condicional de probabilidades*. En aquellas condiciones cuando aparece, precisamente, el suceso B, el comportamiento probabilístico de la magnitud aleatoria ξ se describe por las probabilidades condicionales $\mathbf{P}\{x' \leqslant \xi \leqslant x'' \mid B\}$, hablando en general, se diferencian de las probabilidades iniciales $\mathbf{P}\{x' \leqslant \xi \leqslant x''\}$.

Ejemplo. Imaginémonos que sólo se puede observar una de las dos pruebas Ω_1 o Ω_2, en cada una de las cuales se examina una determinada magnitud aleatoria (ξ_1 en Ω_1 y ξ_2 en Ω_2). Con objeto de elegir una de las pruebas Ω_1 o Ω_2, el observador echa a la suerte, complementariamente, de tal modo que se elige a Ω_1 con la probabilidad dada q_1 y se elige a Ω_2 con la probabilidad q_2 $(q_1 + q_2 = 1)$. En total, en lugar de las magnitudes ξ_1 y ξ_2 aparece la magnitud aleatoria de la forma $\xi = \eta\xi_1 + (1 - \eta)\,\xi_2$, donde η no depende de ξ_1, ξ_2 y toma con las probabilidades q_1, q_2 los valores 1 ó 0, respectivamente. Se ve fácilmente, que para la condición $\eta = 1$, esta magnitud ξ coincide con ξ_1 y para la condición $\eta = 0$, con la magnitud ξ_2. De acuerdo con esto, la distribución de probabilidades de la magnitud aleatoria ξ es tal que

$$\mathbf{P}\{x' \leqslant \xi \leqslant x'' \mid \eta = 1\} = \mathbf{P}\{x' \leqslant \xi_1 \leqslant x''\},$$

$$\mathbf{P}\{x' \leqslant \xi \leqslant x'' \mid \eta = 0\} = \mathbf{P}\{x' \leqslant \xi_2 \leqslant x''\},$$

y según la fórmula de la probabilidad total

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = q_1 \cdot \mathbf{P}\{x' \leqslant \xi_1 \leqslant x''\} + q_2 \cdot \mathbf{P}\{x' \leqslant \xi_2 \leqslant x''\}.$$

Si decimos que ξ_1 es una magnitud discreta que toma distintos valores de x con las correspondientes probabilidades $P_{\xi_1}(x)$, y ξ_2 es una magnitud distribuida continuamente con una densidad de probabilidad $p_{\xi_2}(x)$, entonces

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = q_1 \cdot \sum_{x'}^{x''} P_{\xi_1}(x) + q_2 \cdot \int_{x'}^{x''} p_{\xi_2}(x)\,dx.$$

Si ξ y η son magnitudes aleatorias discretas, entonces para el suceso B del tipo $\{\eta = y\}$ la magnitud ξ ligada con η ya no tendrá la distribución de probabilidades anterior $P_\xi(x)$, sino la tal llamada distribución condicional $P_\xi(x \mid y)$ que de acuerdo con la fórmula (2.4) se determina así:

$$P_\xi(x \mid y) = \frac{P_{\xi,\eta}(x,\ y)}{P_\eta(y)} = \frac{P_{\xi,\eta}(x,\ y)}{\sum\limits_x P_{\xi,\eta}(x,\ y)}, \quad -\infty < x < \infty, \qquad (3.16)$$

donde x recorre todos los valores posibles de la magnitud discreta ξ, y $P_{\xi,\eta}(x,\ y)$ es la distribución conjunta de probabilidades de ξ y η.

Análogamente, si ξ y η tienen una densidad de distribución conjunta $p_{\xi,\eta}(x,\ y)$, entonces consideraremos que para el valor fijado $\eta = y$, la magnitud ξ tiene *una densidad de distribución condicional* $p_\xi(x \mid y)$, que determinamos por la fórmula

$$p_\xi(x \mid y) = \frac{p_{\xi,\eta}(x,\ y)}{p_\eta(y)} = \frac{p_{\xi,\eta}(x,\ y)}{\int\limits_{-\infty}^{\infty} p_{\xi,\eta}(x,\ y)\,dx}, \quad -\infty < x < \infty. \qquad (3.17)$$

De las fórmulas (3.16) y (3.17) se ve, directamente, que la magnitud aleatoria ξ no depende de η, cuando, y sólo cuando su distribución de probabilidades condicional (para la condición $\eta = y$) no depende de y y coincide con la distribución inicial (incondicional).

Señalemos las siguientes igualdades (compárese con la fórmula de la probabilidad total):

$$P_\xi(x) = \sum_y P_\xi(x \mid y)\, P_\eta(y) \qquad (3.18)$$

para las magnitudes discretas ξ y η,

$$p_\xi(x) = \int\limits_{-\infty}^{\infty} p_\xi(x \mid y)\, p_\eta(y)\,dy \qquad (3.19)$$

para las magnitudes ξ y η con densidades de distribución de probabilidades $p_\xi(x)$ y $p_\eta(y)$.

Para ilustrar, cuán útil puede ser la consideración de tal o cual distribución condicional, nos dirigimos de nuevo a la suma de magnitudes aleatorias independientes ξ_1, ξ_2 con densidad de probabilidad $p_1(x)$, $p_2(x)$. Hallemos la distribución de la magnitud $\eta = \xi_1 + \xi_2$.

Señalemos, primeramente, que la distribución condicional de probabilidades de la magnitud η en condiciones, cuando $\xi_2 = x$, coincide con la distribución de la magnitud $\xi_1 + x$, que tiene una densidad $p_1(y - x)$, $-\infty < y < \infty$ (recordemos, que ξ_1 no depende de ξ_2, de modo que la densidad condicional de distribución de la magnitud ξ_1 coincide con la densidad incondicional $p_1(y)$). Según

la fórmula general (3.19) obtenemos la expresión ya conocida (3.15):

$$p_\eta(y) = \int_{-\infty}^{\infty} p_1(y-x)\, p_2(x)\, dx.$$

5. Magnitudes aleatorias multidimensionales. El conjunto de magnitudes aleatorias $\xi_1, \ldots, \xi_n$ (que toman los valores numéricos $x_1, \ldots, x_n$) suele ser cómodo examinarlo como una magnitud aleatoria vectorial $\xi = (\xi_1, \ldots, \xi_n)$ con las coordenadas $\xi_1, \ldots, \xi_n$, que en dependencia del caso toma tales o cuales valores $x = (x_1, \ldots, x_n)$ en el espacio vectorial E^n de n dimensiones.

Para las magnitudes discretas, la probabilidad de caer el punto ξ en tal o cual zona $A \subseteq E^n$ se determina así:

$$\mathbf{P}\{\xi \in A\} = \sum_{x \in A} P_\xi(x), \tag{3.20}$$

donde las probabilidades

$$P_\xi(x) = \mathbf{P}\{\xi = x\} = \mathbf{P}\{\xi_1 = x_1, \ldots, \xi_n = x_n\} = P_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) \tag{3.21}$$

dan *la distribución conjunta de probabilidades* de las magnitudes $\xi_1, \ldots, \xi_n$. Para las magnitudes distribuidas continuamente $\xi_1, \ldots, \xi_n$ con la densidad de probabilidad

$$p_\xi(x) = p_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$$

$$\left(p_\xi(x) \geqslant 0, \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p_\xi(x)\, dx_1, \ldots, dx_n = 1\right)$$

la probabilidad del suceso $\{\xi \in A\}$ se determina así:

$$\mathbf{P}\{\xi \in A\} = \int_A \ldots \int p_\xi(x)\, dx. \tag{3.22}$$

La diferencia con la fórmula (3.9) expuesta anteriormente, aquí sólo consiste en qué en lugar de la variable $x = (x_1, x_2)$ en el plano E^2 tenemos $x = (x_1, \ldots, x_n) \in E^n$. Todo lo dicho anteriormente, sobre las magnitudes unidimensionales con cambios análogos se extiende a las magnitudes multidimensionales. Particularmente, de una forma evidente se extienden al caso de n variables $(x_1, \ldots, x_n)$ las fórmulas (3.7), (3.10) y también las fórmulas de transformación (1.4), (1.7) y (3.13).

La independencia de las magnitudes vectoriales ξ_1 y ξ_2 significa las mismas propiedades (expresadas por las relaciones (3.11), (3.12)), que para las magnitudes corrientes. Sin ningún cambio se determinan también las distribuciones condicionales de probabilidades de la

magnitud vectorial ξ en relación a la magnitud también vectorial η (véase (3.16), (3.17)).

El único concepto, esencialmente nuevo, para el conjunto de las magnitudes aleatorias $\xi_1, \xi_2, \ldots$ es su *independencia* mutua (abreviadamente: *independencia*). Precisamente, las magnitudes aleatorias (unidimensionales) $\xi_1, \xi_2 \ldots$ se llaman mutuamente independientes, si para cualesquiera x_1', x_1''; x_2', x_2''; ... son mutuamente independientes los sucesos

$$\{x_1' \leqslant \xi_1 \leqslant x_1''\}, \quad \{x_2' \leqslant \xi_2 \leqslant x_2''\}, \ldots$$

Para las magnitudes discretas $\xi_1, \ldots, \xi_n$ la independencia significa que

$$P_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = P_{\xi_1}(x_1) \ldots P_{\xi_n}(x_n), \tag{3.23}$$

y para las magnitudes distribuidas continuamente con las densidades de probabilidad $p_{\xi_1}(x_1), \ldots, p_{\xi_n}(x_n)$, significa que su distribución conjunta de probabilidades tiene la densidad

$$p_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = p_{\xi_1}(x_1) \ldots p_{\xi_n}(x_n). \tag{3.24}$$

§ 4. ESPERANZAS MATEMATICAS DE LAS MAGNITUDES ALEATORIAS

1. Esperanza matemática, definición y algunas fórmulas. Sea ξ una magnitud aleatoria discreta, que toma cada valor posible de x con la correspondiente probabilidad $P_\xi(x)$. Se dice que la magnitud aleatoria ξ tiene *una esperanza matemática* finita $\mathbf{M}\xi$:

$$\mathbf{M}\xi = \sum_{-\infty}^{\infty} x P_\xi(x) \tag{4.1}$$

(llamada también *valor medio* de esta magnitud aleatoria), si la serie en (4.1) converge absolutamente ($\sum_{-\infty}^{\infty} |x| P_\xi(x) < \infty$).

Ejemplo. El valor medio de la magnitud aleatoria ξ, que toma uno de los N valores posibles $x = x_1, \ldots, x_N$ con una misma probabilidad $\left(\text{igual a } \frac{1}{N}\right)$, es

$$\mathbf{M}\xi = \frac{1}{N} \sum_{k=1}^{N} x_k.$$

Examinemos la magnitud aleatoria $\eta = \varphi(\xi)$ que tiene esperanza matemática, donde $y = \varphi(x)$ es una función determinada de la variable x. *Para el valor medio* $\mathbf{M}\varphi(\xi)$ *tiene lugar la fórmula*

$$\mathbf{M}\varphi(\xi) = \sum_{-\infty}^{\infty} \varphi(x) P_\xi(x), \tag{4.2}$$

donde la sumación se efectúa por todos los valores de x, para los cuales las probabilidades $P_\xi(x) > 0$.

En efecto, la magnitud discreta $\eta = \varphi(\xi)$ puede tomar, solamente, los valores $y = \varphi(x)$, donde x recorre los valores posibles de la magnitud discreta ξ, siendo $P_\eta(y) = \sum_{x:\varphi(x)=y} P_\xi(x)$, se ve fácilmente que si la serie $\sum_{-\infty}^{\infty} yP_\eta(y)$ converge absolutamente, entonces

$$\mathbf{M}\eta = \sum_{-\infty}^{\infty} yP_\eta(y) = \sum_y y \sum_{x:\varphi(x)=y} P_\xi(x) = \sum_{-\infty}^{\infty} \varphi(x) P_\xi(x).$$

Exactamente igual, para el valor medio de la magnitud aleatoria $\eta = \varphi(\xi_1, \xi_2)$, que es una determinada función de las magnitudes ξ_1, ξ_2 con la distribución conjunta de probabilidades $P_{\xi_1,\xi_2}(x_1, x_2)$, tenemos

$$\mathbf{M}\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \varphi(x_1, x_2) P_{\xi_1,\xi_2}(x_1, x_2) \tag{4.3}$$

(formalmente (4.3) coincide con (4.2), si suponemos que

$$\xi = (\xi_1, \xi_2) \quad \text{y} \quad x = (x_1, x_2)).$$

Según la fórmula (4.2), la esperanza matemática de la magnitud $|\xi|$ es $\mathbf{M}|\xi| = \sum_{-\infty}^{\infty} |x| P_\xi(x)$, de modo que la condición de la convergencia absoluta de la serie $\sum_{-\infty}^{\infty} xP_\xi(x)$ puede ser expresada de la siguiente forma: $\mathbf{M}|\xi| < \infty$ (señalemos que la expresión $\mathbf{M}|\xi| = \sum_{-\infty}^{\infty} |x| P_\xi(x)$ tiene sentido para cualesquiera magnitudes, solamente, puede ocurrir que $\mathbf{M}|\xi| = \infty$).

La esperanza matemática posee las siguientes propiedades.

El valor medio de una magnitud constante es igual a sí misma y en particular.

$$\mathbf{M}1 = 1.$$

Si existe la esperanza matemática $\mathbf{M}\xi$, entonces para cualquier factor constante k, tendremos para la magnitud $k \cdot \xi$ que

$$\mathbf{M}(k \cdot \xi) = k \cdot \mathbf{M}\xi.$$

Para cualesquiera magnitudes ξ_1 y ξ_2 (que tienen las esperanzas matemáticas $M\xi_1$ y $M\xi_2$)

$$M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2; \tag{4.4}$$

si $\xi_1 \leqslant \xi_2$, entonces

$$M\xi_1 \leqslant M\xi_2; \tag{4.5}$$

si las magnitudes aleatorias ξ_1 y ξ_2 son independientes, entonces

$$M(\xi_1 \cdot \xi_2) = M\xi_1 \cdot M\xi_2. \tag{4.6}$$

Para cerciorarse de la justeza de las relaciones indicadas utilizemos las fórmulas (4.2), (4.3). Para $\varphi(x) = k \cdot x$ tenemos

$$M\varphi(\xi) = \sum_{-\infty}^{\infty} k \cdot x P_\xi(x) - k \cdot \sum_{-\infty}^{\infty} x P_\xi(x) = kM\xi.$$

Tomando $\varphi(x_1, x_2) = x_1 + x_2$, obtenemos

$$M\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} x_1 \sum_{-\infty}^{\infty} P_{\xi_1, \xi_2}(x_1, x_2) + \sum_{-\infty}^{\infty} x_2 \sum_{-\infty}^{\infty} P_{\xi_1, \xi_2}(x_1, x_2),$$

de donde (véase también (3.7)) se ve, que la igualdad (4.4) en efecto tiene lugar. Para cualquier función no negativa $\varphi(x_1, x_2)$ en todos los valores de x_1, x_2 para los cuales las probabilidades $P_{\xi_1, \xi_2}(x_1, x_2) > 0$,

$$M\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} \varphi(x_1, x_2) P_{\xi_1, \xi_2}(x_1, x_2) \geqslant 0,$$

y, en particular, $M(\xi_2 - \xi_1) \geqslant 0$ para $\xi_2 - \xi_1 \geqslant 0$, de donde se deduce que en este caso $M\xi_1 \leqslant M\xi_2$. Para las magnitudes independientes ξ_1 y ξ_2 la distribución conjunta de probabilidades se determina por la fórmula (3.11) y para $\varphi(x_1, x_2) = x_1 \cdot x_2$ tenemos

$$M\varphi(\xi_1, \xi_2) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} (x_1 \cdot x_2) P_{\xi_1}(x_1) P_{\xi_2}(x_2) =$$

$$= \sum_{-\infty}^{\infty} x_1 P_{\xi_1}(x_1) \cdot \sum_{-\infty}^{\infty} x_2 P_\xi(x_2) = M\xi_1 \cdot M\xi_2.$$

La fórmula (4.4) se extiende por inducción a cualquier número de magnitudes aleatorias $\xi_1, \ldots, \xi_n$ que tienen los valores medios:

$$M(\xi_1 + \ldots + \xi_n) = M\xi_1 + \ldots + M\xi_n;$$

también se puede decir lo mismo sobre la fórmula (4.6); para cualesquiera magnitudes mutuamente independientes $\xi_1, \ldots, \xi_n$

$$M(\xi_1 \ldots \xi_n) = M\xi_1 \ldots M\xi_n.$$

Volvamos a la magnitud aleatoria arbitraria ξ (no obligatoriamente discreta). La definición de la esperanza matemática $\mathbf{M}\xi$ está basada en que cualquier magnitud ξ se puede aproximar con magnitudes discretas, tan exactamente, como se desee. Por ejemplo, si dividimos la recta real $-\infty < x < \infty$ con los puntos $x_{k,n}$ $(-\infty < k < \infty$, $\sup_k | x_{k,n} - x_{k-1,n} | = \varepsilon_n)$ y determinamos las magnitudes discretas aleatorias ξ_n así:

$$\xi_n = x_{k,n} \quad \text{para} \quad x_{k-1,n} < \xi \leqslant x_{k,n},$$

entonces, evidentemente $| \xi_n - \xi | \leqslant \varepsilon_n$ y para $\varepsilon_n \to 0$ la sucesión de magnitudes discretas ξ_n, $n = 1, 2, \ldots$, da la aproximación tan exacta como se desee para la magnitud inicial ξ. Si las magnitudes ξ_n tienen las esperanzas matemáticas $\mathbf{M}\xi_n$ (es decir, $\mathbf{M} | \xi_n | < \infty$), entonces, según la desigualdad general (4.5)

$$| \mathbf{M}\xi_n - \mathbf{M}\xi_m | \leqslant \mathbf{M} | \xi_n - \xi_m | \leqslant \varepsilon_n + \varepsilon_m \to 0$$
$$(| \xi_n - \xi_m | \leqslant | \xi_n - \xi | + | \xi_m - \xi | \leqslant \varepsilon_n + \varepsilon_m)$$

y, por consiguiente, existe el límite $\lim_{n\to\infty} \mathbf{M}\xi_n$. Este límite se llama *esperanza matemática* (o *valor medio*) de la magnitud aleatoria ξ:

$$\mathbf{M}\xi = \lim_{n\to\infty} \sum_{-\infty}^{\infty} x_{k,n} \mathbf{P}(x_{k-1,n} < \xi \leqslant x_{k,n}). \tag{4.7}$$

(Se ve, fácilmente, que para cualquier sucesión de magnitudes ξ_n, $n = 1, 2, \ldots$, del tipo indicado, o en otras palabras, para cualquier división de la recta real en intervalos $(x_{k-1,n}, x_{k,n}]$, para $\sup_k | x_{k,n} - x_{k-1,n} | \to 0$, el valor límite $\mathbf{M}\xi$ en (4.7) será el mismo).

Evidentemente, las relaciones (4.4)—(4.6) se conservan durante el paso límite utilizado por nosotros desde las magnitudes discretas hasta las magnitudes aleatorias arbitrarias.

Demostremos que *para la magnitud aleatoria* ξ *con la densidad de probabilidad* $p_\xi(x)$ *la esperanza matemática será*

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} x p_\xi(x)\, dx \tag{4.8}$$

$\left(\text{donde la integral converge absolutamente: } \int_{-\infty}^{\infty} |x| p_\xi(x)\, dx < \infty\right)$.

En efecto, suponiendo $\varphi(x) = x$ y $\varphi_n(x) = x_{k,n}$ para $x_{k-1,n} < x \leqslant x_{k,n}$ tenemos

$$\mathbf{M}\xi_n = \sum_{-\infty}^{\infty} \int_{x_{k-1,n}}^{x_{k,n}} x_{k,n} p_\xi(x)\, dx = \int_{-\infty}^{\infty} \varphi_n(x) p_\xi(x)\, dx,$$

donde

$$\left|\int_{-\infty}^{\infty}\varphi_n(x)\,p_\xi(x)\,dx-\int_{-\infty}^{\infty}\varphi(x)\,p_\xi(x)\,dx\right|\leqslant\int_{-\infty}^{\infty}|\varphi_n(x)-\varphi(x)|\times$$

$$\times\,p_\xi(x)\,dx\leqslant\varepsilon_n\int_{-\infty}^{\infty}p_\xi(x)\,dx=\varepsilon_n\to 0 \text{ para } n\to\infty$$

y, por consiguiente, $\mathbf{M}\xi_n\to\int_{-\infty}^{\infty}\varphi(x)\,p_\xi(x)\,dx$, es decir $\mathbf{M}\xi=$ $=\int_{-\infty}^{\infty}xp_\xi(x)\,dx$. (Se ve fácilmente, que la condición $\mathbf{M}\,|\,\xi_n\,|=$ $=\int_{-\infty}^{\infty}|\,\varphi_n(x)\,|\,p_\xi(x)\,dx<\infty$ presupuesta en la definición de la esperanza matemática $\mathbf{M}\xi$ es equivalente a que $\int_{-\infty}^{\infty}|\,\varphi(x)\,|\,p_\xi(x)\,dx<\infty$, $\varphi(x)=x$.)

Utilizando un método análogo de aproximación en la función $y=\varphi(x)$ con funciones constantes a trozos $\varphi_n(x)$, para la magnitud aleatoria $\eta=\varphi(\xi)$ en condiciones que $\int_{-\infty}^{\infty}|\,\varphi(x)\,|\,p_\xi(x)\,dx<\infty$ se puede deducir que

$$\mathbf{M}\varphi(\xi)=\int_{-\infty}^{\infty}\varphi(x)\,p_\xi(x)\,dx. \tag{4.9}$$

La misma fórmula tiene lugar cuando $\xi=(\xi_1,\ \xi_2)$ para la magnitud aleatoria $\eta=\varphi(\xi_1,\ \xi_2)$: o más exacto,

$$\mathbf{M}\varphi(\xi_1,\ \xi_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(x_1,\ x_2)\,p_{\xi_1,\,\xi_2}(x_1,\ x_2)\,dx_1\,dx_2, \tag{4.10}$$

donde $p_{\xi_1,\xi_2}(x_1,\ x_2)$ es la densidad de probabilidad de las magnitudes ξ_1, ξ_2, con ello, la condición de existencia de la esperanza matemática finita ($\mathbf{M}\,|\,\varphi(\xi_1,\ \xi_2)\,|<\infty$) es equivalente a que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\,\varphi(x_1,\ x_2)\,|\,p_{\xi_1,\,\xi_2}(x_1,\ x_2)\,dx_1\,dx_2<\infty.$$

Ejemplo. Sea ξ la magnitud aleatoria distribuida uniformemente en el segmento $[a,\ b]$, es decir, que tiene una densidad de probabilidad

de la forma

$$p_\xi(x) = \begin{cases} \dfrac{1}{b-a} & \text{para } a \leqslant x \leqslant b, \\ 0 & \text{para } x < a,\ x > b. \end{cases}$$

Entonces su valor medio es

$$\mathbf{M}\xi = \int_a^b x p_\xi(x)\, dx = \frac{a+b}{2}.$$

Examinemos la magnitud aleatoria ξ de «tipo combinado» (véase el ejemplo en la pág. 42): ξ toma un valor determinado de x con la probabilidad correspondiente $P_\xi(x)$ $\left(\sum_{-\infty}^{\infty} P_\xi(x) = q_1\right)$, y en general cae en tal o cual intervalo $[x', x'']$ con la probabilidad

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = \sum_{x'}^{x''} P_\xi(x) + \int_{x'}^{x''} p_\xi(x)\, dx,$$

donde $p_\xi(x)$ es una función no negativa, $\int_{-\infty}^{\infty} p_\xi(x)\, dx = q_2$, $q_1 + q_2 = 1$, de modo que las probabilidades $P_\xi(x) = q_1^{-1} \cdot P_\xi(x)$, $-\infty < x < \infty$, y la densidad $p_{\xi_2}(x) = q_2^{-1} \cdot p_\xi(x)$, $-\infty < x < \infty$, dan, correspondientemente, las distribuciones de probabilidades de la magnitud discreta, digamos ξ_1 y la magnitud ξ_2 distribuida continuamente. Para tal magnitud ξ, la esperanza matemática $\mathbf{M}\xi$ puede ser expresada por la fórmula (compárese con (4.1), (4.8))

$$\mathbf{M}\xi = q_1 \mathbf{M}\xi_1 + q_2 \mathbf{M}\xi_2 = \sum_{-\infty}^{\infty} x P_\xi(x) + \int_{-\infty}^{\infty} x p_\xi(x)\, dx. \qquad (4.11)$$

El concepto de esperanza matemática $\mathbf{M}\xi$ también se extiende a las magnitudes aleatorias ξ, que toman valores complejos: $\xi = \xi_1 + i\xi_2$, donde ξ_1 y ξ_2 son magnitudes aleatorias reales. Precisamente, se supone que

$$\mathbf{M}\xi = \mathbf{M}\xi_1 + i\mathbf{M}\xi_2.$$

Se ve, fácilmente, que todas las propiedades citadas de las esperanzas matemáticas también tienen lugar para las magnitudes complejas de la forma $\xi = \xi_1 + i\xi_2$ (y en general, para las funciones de valores complejos $\varphi(\xi_1, \xi_2)$; véanse en particular, las fórmulas (4.3) y (4.10)).

Como conclusión de este apartado observemos, que las fórmulas, anteriormente indicadas, (4.2), (4.3) y (4.9), (4.10), de forma evidente

se extienden a las magnitudes aleatorias multidimensionales $\xi = (\xi_1, \ldots, \xi_n)$.

2. Momentos, dispersión y desigualdad de Chebishev. Se llama *momento* (de orden k) de la magnitud aleatoria ξ al valor medio $\mathbf{M}\xi^k$. De acuerdo a las fórmulas generales (4.2) y (4.9)

$$\mathbf{M}\xi^k = \sum_{-\infty}^{\infty} x^k P_\xi(x)$$

para las magnitudes discretas y

$$\mathbf{M}\xi^k = \int_{-\infty}^{\infty} x^k p_\xi(x)\, dx$$

para las magnitudes distribuidas continuamente.

Se llama segundo momento $\mathbf{M}\xi^2$ al *valor de la media cuadrática* de la magnitud aleatoria ξ.

Es justa la siguiente *desigualdad (Cauchy—Bunjakovsky)*:

$$\mathbf{M} \mid \xi_1 \cdot \xi_2 \mid \leqslant \sqrt{\mathbf{M}\xi_1^2} \cdot \sqrt{\mathbf{M}\xi_2^2} \tag{4.12}$$

para cualesquiera magnitudes ξ_1, ξ_2 (que tienen un segundo momento finito).

En realidad, ya que $\mid \xi_1 \cdot \xi_2 \mid \leqslant 1/2\,(\xi_1^2 + \xi_2^2)$, de la limitación de los segundos momentos $\mathbf{M}\xi_1^2$ y $\mathbf{M}\xi_2^2$ se deduce, que $\mathbf{M} \mid \xi_1 \cdot \xi_2 \mid < \infty$. Luego, la forma cuadrática determinada positivamente

$$\mathbf{M}\,(x_1 \mid \xi_1 \mid + x_2 \mid \xi_2 \mid)^2 = x_1^2 \cdot \mathbf{M}\xi_1^2 + 2x_1x_2 \cdot \mathbf{M} \mid \xi_1\xi_2 \mid + x_2^2 \cdot \mathbf{M}\xi_2^2$$

de las variables x_1, x_2 tiene un determinante no negativo, igual a $\mathbf{M}\xi_1^2 \cdot \mathbf{M}\xi_2^2 - (\mathbf{M} \mid \xi_1 \cdot \xi_2 \mid)^2$, lo que da la desigualdad (4.12).

Suponiendo que $\xi_1 = \xi$, $\xi_2 = 1$, obtenemos de la fórmula (4.12) que

$$\mathbf{M} \mid \xi \mid \leqslant \sqrt{\mathbf{M}\xi^2}. \tag{4.13}$$

De (4.12) también se deduce la siguiente desigualdad importante

$$\sqrt{\mathbf{M} \mid \xi_1 + \xi_2 \mid^2} \leqslant \sqrt{\mathbf{M}\xi_1^2} + \sqrt{\mathbf{M}\xi_2^2}. \tag{4.14}$$

En efecto,

$$\mathbf{M} \mid \xi_1 + \xi_2 \mid^2 = \mathbf{M}\xi_1^2 + 2\mathbf{M}\xi_1\xi_2 + \mathbf{M}\xi_2^2 \leqslant \mathbf{M}\xi_1^2 + 2\sqrt{\mathbf{M}\xi_1^2}\sqrt{\mathbf{M}\xi_2^2} + \mathbf{M}\xi_2^2 = (\sqrt{\mathbf{M}\xi_1^2} + \sqrt{\mathbf{M}\xi_2^2})^2.$$

Luego, tiene lugar la siguiente *desigualdad de Chebishev*:

$$\mathbf{P}\{\mid \xi \mid > \varepsilon\} \leqslant \frac{\mathbf{M} \mid \xi \mid^2}{\varepsilon^2}, \tag{4.15}$$

para cualquiera $\varepsilon > 0$.

Esta se puede deducir fácilmente, de la desigualdad general (4.5). En efecto, *si suponemos que*

$$\xi_1 = \begin{cases} 0 & \text{para } |\xi| \leqslant \varepsilon, \\ \varepsilon & \text{para } |\xi| > \varepsilon, \end{cases}$$

entonces, evidentemente, $\xi_1 \leqslant |\xi|$ y

$$\mathbf{M}|\xi|^2 \geqslant \mathbf{M}\xi_1^2 = \varepsilon^2 \cdot \mathbf{P}\{|\xi| > \varepsilon\},$$

que es equivalente a la relación (4.15).

La desigualdad de Chebishev muestra que si el valor de la media cuadrática $\mathbf{M}\xi^2$ es pequeño en relación a ε^2, digamos $\mathbf{M}\xi^2/\varepsilon^2 \leqslant \delta$, y se puede despreciar, prácticamente, la posibilidad de realización del suceso $\{|\xi| > \varepsilon\}$ de pequeña probabilidad δ, entonces también será pequeña la propia magnitud aleatoria ξ, a saber: $|\xi| \leqslant \varepsilon$; en particular, *si* $\mathbf{M}|\xi|^2 = 0$, entonces $\xi = 0$ *con la probabilidad* 1 (en general, *si* $\xi \geqslant 0$ y $\mathbf{M}\xi = \mathbf{M}|\sqrt{\xi}|^2 = 0$, *entonces* $\xi = 0$ con la probabilidad 1).

Examinemos la diferencia $\xi - a$, donde $a = \mathbf{M}\xi$ es la esperanza matemática de la magnitud ξ; se llama *dispersión* de la magnitud aleatoria ξ, y se designa por $\mathbf{D}\xi$ al valor medio cuadrático de la diferencia $\xi - a$, es decir

$$\mathbf{D}\xi = \mathbf{M}(\xi - a)^2.$$

Según la desigualdad de Chebishev,

$$\mathbf{P}\{|\xi - a| > \varepsilon\} \leqslant \frac{\mathbf{D}\xi}{\varepsilon^2},$$

de modo que para una dispersión relativamente pequeña $\mathbf{D}\xi$, la magnitud aleatoria ξ toma con una gran probabilidad un valor cercano al a (en este sentido también se diferencia poco de la constante a):

$$\mathbf{P}\{|\xi - a| \leqslant \varepsilon\} \geqslant 1 - \frac{\mathbf{D}\xi}{\varepsilon^2},$$

en particular, si $\mathbf{D}\xi = 0$, entonces $\xi = a$ *con la probabilidad* 1 (evidentemente, la dispersión de la magnitud constante $\xi = a$ será siempre igual a 0).

Señalemos las siguientes propiedades de la dispersión. Tiene lugar la fórmula

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - a^2 \quad (a = \mathbf{M}\xi).$$

(En realidad, $\mathbf{M}(\xi - a)^2 = \mathbf{M}\xi^2 - 2a \cdot \mathbf{M}\xi + a^2 = \mathbf{M}\xi^2 - a^2$).

Para cualquier factor constante k, tenemos para la magnitud aleatoria $k \cdot \xi$ que

$$\mathbf{D}(k \cdot \xi) = k^2 \cdot \mathbf{D}\xi.$$

Sean ξ_1 y ξ_2 magnitudes aleatorias independientes. Entonces

$$\mathbf{D}(\xi_1 + \xi_2) = \mathbf{D}\xi_1 + \mathbf{D}\xi_2. \tag{4.16}$$

En efecto, si suponemos que $a_1 = \mathbf{M}\xi_1$ y $a_2 = \mathbf{M}\xi_2$, entonces, según la relación (4.6) tendremos

$$\mathbf{M}(\xi_1 - a_1)(\xi_2 - a_2) = \mathbf{M}(\xi_1 - a_1) \cdot \mathbf{M}(\xi_2 - a_2) = 0$$

y

$$\mathbf{D}(\xi_1 + \xi_2) = \mathbf{M}[(\xi_1 - a_1) + (\xi_2 - a_2)]^2 =$$
$$= \mathbf{M}(\xi_1 - a_1)^2 + 2\mathbf{M}(\xi_1 - a_1)(\xi_2 - a_2) +$$
$$+ \mathbf{M}(\xi_2 - a_2)^2 = \mathbf{M}(\xi_1 - a_1)^2 + \mathbf{M}(\xi_2 - a_2)^2 = \mathbf{D}\xi_1 + \mathbf{D}\xi_2.$$

Evidentemente, la fórmula (4.16) se extiende a cualquier número de magnitudes aleatorias mutuamente independientes $\xi_1, \ldots, \xi_n$:

$$\mathbf{D}(\xi_1 + \ldots + \xi_n) = \mathbf{D}\xi_1 + \ldots + \mathbf{D}\xi_n.$$

3. *Esperanza matemática condicional.* *Si* es conocido, que aparece el suceso determinado B, el comportamiento probable de la magnitud aleatoria ξ (ligada de algún modo al suceso B) se describe por las probabilidades condicionales $\mathbf{P}\{x' \leqslant \xi \leqslant x'' \mid B\}$. El valor medio de la magnitud ξ en relación a tal distribución condicional de probabilidades se llama *esperanza matemática condicional* (*o valor medio condicional*) y se designa así: $\mathbf{M}(\xi \mid B)$.

Por ejemplo, para la magnitud aleatoria $\xi = \eta\xi_1 + (1 - \eta)\xi_2$, donde η no depende de ξ_1, ξ_2 y toma los valores correspondientes 1 ó 0 con las probabilidades q_1, q_2 ($q_1 + q_2 = 1$) (véase el ejemplo de *la pág. 42*), *su esperanza matemática condicional*, al realizarse el suceso $\{\eta = 1\}$ (en este caso ξ coincide con ξ_1), es $\mathbf{M}(\xi \mid \eta = 1) = \mathbf{M}\xi_1$ y, al ocurrir el suceso $\{\eta = 0\}$ (en este caso ξ coincide con ξ_2), es $\mathbf{M}(\xi \mid \eta = 1) = \mathbf{M}\xi_2$.

Para la magnitud discreta ξ con la distribución condicional $P_\xi(x \mid B) = \mathbf{P}\{\xi = x \mid B\}$ *la esperanza matemática condicional*, según la fórmula (4.1), es

$$\mathbf{M}(\xi \mid B) = \sum_{-\infty}^{\infty} x P_\xi(x \mid B).$$

En particular, si se trata del suceso B del tipo $\{\eta = y\}$, entonces, para las magnitudes discretas ξ y η el valor medio correspondiente es

$$\mathbf{M}(\xi \mid y) = \sum_{-\infty}^{\infty} x P_\xi(x \mid y),$$

donde $P_\xi(x \mid y)$ se da por la fórmula (3.16). El valor medio condicional de la magnitud ξ para las magnitudes distribuidas continuamente,

con la condición $\{\eta = y\}$, será (compárese con (4.8))

$$\mathbf{M}(\xi \mid y) = \int_{-\infty}^{\infty} x p_{\xi}(x \mid y)\, dx,$$

donde la densidad condicional de probabilidad $p_{\xi}(x \mid y)$ se da por la fórmula (3.17). Análogamente se determina el valor medio condicional $\mathbf{M}(\xi \mid y_1, \ldots, y_n)$ en relación a determinadas magnitudes $\eta_1, \ldots, \eta_n$ (con la condición de que $\eta_1 = y_1, \ldots, \eta_n = y_n$). La distribución condicional de probabilidades correspondiente se da para las magnitudes discretas $\xi, \eta_1, \ldots, \eta_n$ por las probabilidades

$$P_{\xi}(x \mid y_1, \ldots, y_n) = \frac{P_{\xi, \eta_1, \ldots, \eta_n}(x, y_1, \ldots, y_n)}{P_{\eta_1, \ldots, \eta_n}(y_1, \ldots, y_n)},$$

y para las magnitudes distribuidas continuamente por la densidad de probabilidad

$$p_{\xi}(x \mid y_1, \ldots, y_n) = \frac{p_{\xi, \eta_1, \ldots, \eta_n}(x, y_1, \ldots, y_n)}{p_{\eta_1, \ldots, \eta_n}(y_1, \ldots, y_n)}.$$

En adelante para la unificación de las designaciones en el caso multidimensional, en lugar de $(y_1, \ldots, y_n)$ simplemente escribiremos y. También con este fin, en lugar del suceso B examinaremos la magnitud η, que será igual a 1 cuando aparezca B e igual a 0 en caso contrario; entonces $\mathbf{M}(\xi \mid B) = \mathbf{M}(\xi \mid y)$, $y = 1$.

Evidentemente, para un valor fijado de y la esperanza matemática condicional $\mathbf{M}(\xi \mid y)$ posee todas las propiedades de las esperanzas matemáticas establecidas anteriormente. En particular,

$$\mathbf{M}(k \cdot \xi \mid y) = k \cdot \mathbf{M}(\xi \mid y)$$

para cualquier constante k,

$$\mathbf{M}(\xi_1 + \xi_2 \mid y) = \mathbf{M}(\xi_1 \mid y) + \mathbf{M}(\xi_2 \mid y),$$

etc. (véanse (4.2)—(4.11)). Señalemos especialmente las siguientes propiedades:

a) *si* ξ *no depende de la magnitud* η, *entonces*

$$\mathbf{M}(\xi \mid y) \equiv \mathbf{M}\xi; \tag{4.17}$$

b) *si* ξ_1 *no depende del par de magnitudes* (ξ_2, η), entonces

$$\mathbf{M}(\xi_1 \cdot \xi_2 \mid y) = \mathbf{M}(\xi_1) \cdot \mathbf{M}(\xi_2 \mid y). \tag{4.18}$$

La igualdad (4.17) es evidente. La igualdad (4.18) se obtiene, exactamente, igual que (4.6), ya que incluso en condiciones $\{\eta = y\}$ las magnitudes ξ_1 y ξ_2 quedan independientes, su distribución con-

dicional conjunta es tal que

$$P_{\xi_1, \xi_2}(x_1, x_2 \mid y) = \frac{P_{\xi_1, \xi_2, \eta}(x_1, x_2, y)}{P_\eta(y)} = \\ = \frac{P_{\xi_1}(x_1) \cdot P_{\xi_2, \eta}(x_2, y)}{P_\eta(y)} = P_{\xi_1}(x_1) \cdot P_{\xi_2}(x_2 \mid y)$$

(para las magnitudes discretas) y

$$p_{\xi_1, \xi_2}(x_1, x_2 \mid y) = \frac{p_{\xi_1, \xi_2, \eta}(x_1, x_2, y)}{p_\eta(y)} = \\ = \frac{p_{\xi_1}(x_1) \cdot p_{\xi_2, \eta}(x_2 y)}{p_\eta(y)} = p_{\xi_1}(x_1)\, p_{\xi_2}(x_2 \mid y)$$

(para las magnitudes distribuidas continuamente con una densidad conjunta $p_{\xi_1, \xi_2, \eta}(x_1, x_2, y)$). Naturalmente, la igualdad (4.18) es también justa para la magnitud vectorial $\eta = (\eta_1, \ldots, \eta_n)$, cuando $y = (y_1, \ldots, y_n)$.

Luego señalemos la fórmula de *la esperanza matemática total*, análoga a (2.5):

$$\mathbf{M}\xi = \sum_k \mathbf{M}(\xi \mid B_k) \cdot \mathbf{P}(B_k),$$

donde $B_1, B_2, \ldots$, es el sistema total de sucesos que no se cortan. Del mismo tipo son las siguientes fórmulas:

$$\mathbf{M}\xi = \sum_{-\infty}^{\infty} \mathbf{M}(\xi \mid y)\, P_\eta(y)$$

para la magnitud discreta η con la distribución $P_\eta(y)$;

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} \mathbf{M}(\xi \mid y)\, p_\eta(y)\, dy$$

para la magnitud η que tiene la densidad de probabilidad $p_\eta(y)$ (estas igualdades se deducen, directamente, de la propia definición de la esperanza matemática condicional).

Si consideramos el valor medio condicional $\mathbf{M}(\xi \mid y)$ como función de la variable y, y en lugar de y ponemos el valor de la magnitud aleatoria η, entonces $\mathbf{M}(\xi \mid \eta)$ se puede considerar como una magnitud aleatoria (que es una función determinada de η). Con tal interpretación la fórmula de la esperanza matemática total toma la siguiente forma:

$$\mathbf{M}[\mathbf{M}\{\xi \mid \eta\}] = \mathbf{M}\xi. \tag{4.19}$$

Observemos también la tal llamada fórmula de *la esperanza matemática repetida*[1]) que generaliza la igualdad (4.19) en el caso, cuando

[1]) Véase más detalladamente sobre la cuestión por ejemplo, el libro de I. I. Guijman y A. V. Skorojod, citado anteriormente.

la distribución inicial de probabilidades por sí misma es condicional:

$$\mathbf{M}\{\mathbf{M}[\xi \mid \eta] \mid \zeta\} = \mathbf{M}\{\xi \mid \zeta\},$$

donde la magnitud aleatoria (multidimensional) ζ es una función de η.

Señalemos la igualdad importante siguiente:

$$\mathbf{M}(\varphi(\eta)\cdot\xi \mid \eta) = \varphi(\eta)\cdot\mathbf{M}(\xi \mid \eta), \tag{4.20}$$

donde $\varphi(\eta)$ significa una magnitud aleatoria, que es una función de η.

En efecto, para cada y fijada, el valor de $\varphi(y)$ es una constante, de modo que

$$\mathbf{M}[\varphi(y)\cdot\xi \mid y] = \varphi(y)\cdot\mathbf{M}(\xi \mid y).$$

Para la magnitud aleatoria ξ que es *el indicador del suceso* A,

$$\xi = \begin{cases} 1 \text{ para la aparición de } A, \\ 0 \text{ en caso contrario,} \end{cases}$$

es evidente que $\mathbf{M}\xi = \mathbf{P}(A)$. La esperanza matemática condicional de tal magnitud ξ para el valor fijado de η, no es otra cosa que *la probabilidad condicional del suceso* A:

$$\mathbf{M}(\xi \mid \eta) = \mathbf{P}(A \mid \eta);$$

en este caso la relación (4.19) da la siguiente *fórmula de la probabilidad total*:

$$\mathbf{P}(A) = \mathbf{M}[\mathbf{P}(A \mid \eta)]. \tag{4.21}$$

Hasta ahora no hemos tocado la cuestión sobre si existe la esperanza matemática condicional finita $\mathbf{M}(\xi \mid \eta)$ para tal o cual valor de η. La existencia de $\mathbf{M}(\xi \mid \eta)$ significa que $\mathbf{M}(|\xi| \mid \eta) < \infty$ (véase el punto 1 en que se define la esperanza matemática). Si existe el valor medio incondicional finito $\mathbf{M}|\xi|$ $(\mathbf{M}|\xi| < \infty)$, entonces la fórmula (4.20) de la esperanza matemática total nos da la siguiente relación:

$$\mathbf{M}[\mathbf{M}(|\xi| \mid \eta)] = \mathbf{M}|\xi| < \infty,$$

de donde se deduce que $\mathbf{M}(|\xi| \mid \eta) < \infty$ con la probabilidad 1.

En efecto, la probabilidad de que la magnitud aleatoria η tome un valor, para el cual $\bar{\xi} = \mathbf{M}(|\xi| \mid \eta)$ no sea mayor que N, según la desigualdad de Chebishev es

$$\mathbf{P}\{\bar{\xi} \leqslant N\} = \mathbf{P}\{\sqrt{\bar{\xi}} \leqslant \sqrt{N}\} \geqslant 1 - \frac{\mathbf{M}\bar{\xi}}{N} = 1 - \frac{\mathbf{M}|\xi|}{N},$$

de aquí inmediatamente se deduce, que

$$\mathbf{P}\{\mathbf{M}(|\xi| \mid \eta) < \infty\} = \lim_{N\to\infty} \mathbf{P}\{\mathbf{M}(|\xi| \mid \eta) < N\} = 1.$$

Así pues, *si* $\mathbf{M} \mid \xi \mid < \infty$, *entonces existe la esperanza matemática condicional* $\mathbf{M}(\xi \mid \eta)$ *para cualquier valor de la magnitud aleatoria* η *con la probabilidad* 1.

4. Distancia media cuadrática y coeficiente de correlación. Designemos por H al conjunto de todas las magnitudes aleatorias ξ, para las cuales $\mathbf{M}\xi^2 < \infty$. Como se deduce de la desigualdad (4.14), para las magnitudes arbitrarias $\xi_1, \xi_2 \in H$, en H entra cualquier combinación lineal $\xi = c_1\xi_1 + c_2\xi_2$ y, en particular, la diferencia $\xi_1 - \xi_2$.

Introduzcamos la distancia media cuadrática entre las magnitudes $\xi_1, \xi_2 \in H$, determinándola como

$$\| \xi_1 - \xi_2 \| = \sqrt{\mathbf{M} \mid \xi_1 - \xi_2 \mid^2}.$$

Esta es análoga a la distancia media cuadrática admitida en el análisis funcional, entre las funciones $\varphi_1(x)$ y $\varphi_2(x)$ de la variable x, $a \leqslant x \leqslant b$, que se determina así:

$$\| \varphi_1 - \varphi_2 \| = \left(\int_a^b | \varphi_1(x) - \varphi_2(x) |^2 f(x)\, dx \right)^{1/2},$$

donde $f(x) \geqslant 0$ es una determinada función llamada peso; aún más, para la condición $\int_a^b f(x)\, dx = 1$ tenemos (véase la fórmula (4.9))

$$\| \varphi_1 - \varphi_2 \| = (\mathbf{M} \mid \varphi_1(\xi) - \varphi_2(\xi) \mid^2)^{1/2},$$

donde ξ, $a \leqslant \xi \leqslant b$, es una magnitud aleatoria con la densidad de probabilidad $f(x)$, $a \leqslant x \leqslant b$.

La distancia media cuadrática $\| \xi_1 - \xi_2 \|$ muestra en grado conocido la gran diferencia entre sí de las magnitudes correspondientes ξ_1, ξ_2; en particular, se puede afirmar a base de la desigualdad de Chebishev, que los valores ξ_1, ξ_2 con una probabilidad no menor que $1 - \frac{1}{\varepsilon^2} \| \xi_1 - \xi_2 \|^2$ se diferenciarán en no más de ε:

$$\mathbf{P}\{ | \xi_1 - \xi_2 | \leqslant \varepsilon \} \geqslant 1 - \frac{1}{\varepsilon^2} \| \xi_1 - \xi_2 \|^2.$$

En este sentido, como ya se indicó, la dispersión $\mathbf{D}\xi = \| \xi - a \|^2$ muestra, en los distintos resultados aleatorios ω, cuán grande es la desviación de la magnitud $\xi(\omega)$ desde su valor medio $a = \mathbf{M}\xi$; a propósito, para cualquier constante c_0

$$\begin{aligned} \mathbf{M}(\xi - c_0)^2 &= \mathbf{M}[(\xi - a) + (a - c_0)]^2 = \\ &= \mathbf{M}(\xi - a)^2 + 2(a - c_0) \cdot \mathbf{M}(\xi - a) + (a - c_0)^2 = \\ &= \mathbf{M}(\xi - a)^2 + (a - c_0)^2 \geqslant \mathbf{M}(\xi - a)^2 = \mathbf{D}\xi, \end{aligned}$$

de modo que el valor medio $a = \mathbf{M}\xi$ es aquella constante, de la cual se desvía lo menos posible la magnitud aleatoria ξ en el sentido de

distancia media cuadrática:

$$\min_{c_0} \| \xi - c_0 \| = \| \xi - a \|. \tag{4.22}$$

Examinemos las magnitudes aleatorias ξ_1 y ξ_2. Planteemos la siguiente tarea: hallar una combinación del tipo $\hat{c}_1 + \hat{c}_2\xi_2$ (donde $\hat{c}_1$ y $\hat{c}_2$ son determinadas constantes), que nos dé la mejor aproximación de la magnitud aleatoria ξ_1 en tal sentido que

$$\mathbf{M}(\xi_1 - \hat{c}_1 - \hat{c}_2\xi_2)^2 = \min_{c_1, c_2} \mathbf{M}(\xi_1 - c_1 - c_2\xi_2)^2 \tag{4.23}$$

(el mínimo se toma por todas las constantes c_1 y c_2).

Supongamos que

$$r = \frac{\mathbf{M}(\xi_1 - a_1)(\xi_2 - a_2)}{\sigma_1\sigma_2}, \tag{4.24}$$

donde

$$a_1 = \mathbf{M}\xi_1, \quad \sigma_1^2 = \mathbf{D}\xi_1 \quad \text{y} \quad a_2 = \mathbf{M}\xi_2, \quad \sigma_2^2 = \mathbf{D}\xi_2.$$

Pasemos para comodidad a las magnitudes aleatorias normalizadas

$$\eta_1 = \frac{\xi_1 - a_1}{\sigma_1} \quad \text{y} \quad \eta_2 = \frac{\xi_2 - a_2}{\sigma_2}.$$

Para cualesquiera constantes c_1 y c_2 tenemos

$$\mathbf{M}(\eta_1 - c_1 - c_2\eta_2)^2 = \mathbf{M}[(\eta_1 - r\eta_2) - c_1 + (r - c_2)\eta_2]^2 = \\ = (1 - r^2) + c_1^2 + (r - c_2)^2.$$

Se ve que el mínimo de la expresión $\mathbf{M}(\eta_1 - c_1 - c_2\eta_2)^2$ se alcanza, cuando $c_1 = 0$ y $c_2 = r$:

$$\min_{c_1, c_2} \mathbf{M}(\eta_1 - c_1 - c_2\eta_2)^2 = \| \eta_1 - r\eta_2 \|^2 = 1 - r^2. \tag{4.25}$$

Expresando la diferencia $\eta_1 - r\eta_2$ a través de las magnitudes iniciales ξ_1 y ξ_2 obtenemos

$$\eta_1 - r\eta_2 = \frac{1}{\sigma_1}\left[\xi_1 - a_1 - r\frac{\sigma_1}{\sigma_2}(\xi_2 - a_2)\right],$$

y evidentemente la combinación lineal buscada $\hat{c}_1 + \hat{c}_2\xi_2$ en (4.23) es

$$\hat{c}_1 + \hat{c}_2\xi_2 = a_1 + r\frac{\sigma_1}{\sigma_2}(\xi_2 - a_2). \tag{4.26}$$

Aquí a_1 y a_2 son las esperanzas matemáticas de las magnitudes aleatorias ξ_1 y ξ_2, σ_1^2 y σ_2^2 son sus dispersiones, y la constante r determinada por la igualdad (4.24) es el tal llamado *coeficiente de correlación* de estas magnitudes aleatorias, que es la característica de enlace más elemental de entre ξ_1 y ξ_2.

Como se ve de la fórmula (4.25) el coeficiente de correlación r siempre se halla entre los límites $-1 \leqslant r \leqslant 1$, con ello, si $r = -1$ ó $r = 1$, entonces la magnitud aleatoria ξ_1 es, sencillamente, una combinación lineal de la forma $\xi_1 = \hat{c}_1 + \hat{c}_2\xi_2$.

En efecto, si $r = -1$ ó $r = 1$, entonces, según la fórmula (4.25) el valor de la media cuadrática de la magnitud $\xi_1 - \hat{c}_1 - \hat{c}_2\xi_2$ es

$$\mathbf{M}(\xi_1 - \hat{c}_1 - \hat{c}_2\xi_2)^2 = \sigma_1^2(1 - r^2) = 0$$

y, por consiguiente, $\xi_1 - \hat{c}_1 - \hat{c}_2\xi_2 = 0$ con la probabilidad 1.

Si las magnitudes aleatorias ξ_1 y ξ_2 son independientes, su coeficiente de correlación es igual a cero, ya que

$$\mathbf{M}(\xi_1 - a_1)(\xi_2 - a_2) = \mathbf{M}(\xi_1 - a_1) \cdot \mathbf{M}(\xi_2 - a_2) = 0.$$

Debemos señalar, que la condición $r = 0$, hablando en general, no lleva consigo la independencia de las magnitudes aleatorias ξ_1 y ξ_2.

Sea, por ejemplo, ξ_1 una magnitud distribuida simétricamente, con una densidad $p_1(x)$ tal que $p(-x) = p(x)$, $-\infty < x < \infty$ y sea $\xi_2 = |\xi_1|$. Entonces, a pesar de que la magnitud ξ_2 también es una función de ξ_1, el coeficiente de correlación de las magnitudes ξ_1 y ξ_2 será igual a cero, ya que

$$a_1 = \mathbf{M}\xi_1 = \int_{-\infty}^{\infty} x p_\xi(x)\, dx = 0$$

y como se ve fácilmente,

$$\mathbf{M}(\xi_1 - a_1)(\xi_2 - a_2) = \mathbf{M}(\xi_1 \cdot \xi_2) = \int_{-\infty}^{\infty} x|x| p_\xi(x)\, dx = 0.$$

Examinemos la tarea general sobre la aproximación óptima de la magnitud aleatoria ξ_0 con ayuda de combinaciones lineales $\sum_{k=1}^{n} c_k\xi_k$ de otras magnitudes aleatorias $\xi_1, \ldots, \xi_n$. La tarea consiste en hallar, según las características dadas

$$a_k = \mathbf{M}\xi_k, \quad \sigma_k^2 = \mathbf{M}(\xi_k - a_k)^2,$$

$$r_{kj} = \frac{\mathbf{M}(\xi_k - a_k)(\xi_j - a_j)}{\sigma_k \sigma_j} \quad (k, j = 0, 1, \ldots, n),$$

tales coeficientes $\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_n$ para los cuales

$$\left\| \xi_0 - \left(\hat{c}_0 + \sum_{k=1}^{n} \hat{c}_k\xi_k\right) \right\| = \min_{c_0, \ldots, c_n} \left\| \xi_0 - \left(c_0 + \sum_{k=1}^{n} c_k\xi_k\right) \right\|. \quad (4.27)$$

De la relación general (4.22) obtenida anteriormente se ve, que si han sido hallados los coeficientes correspondientes $\hat{c}_1, \ldots, \hat{c}_n$,

entonces, la constante $\hat{c}_0$ es necesario elegirla del siguiente modo:

$$\hat{c}_0 = \mathbf{M}\left(\xi_0 - \sum_{k=1}^{n} \hat{c}_k \xi_k\right) = a_0 - \sum_{k=1}^{n} \hat{c}_k a_k.$$

Sin limitar la generalidad, se puede considerar $a_k = 0$, $k = 0, 1, \ldots, n$, y entonces también $\hat{c}_0 = 0$.

Así pues, examinemos las magnitudes aleatorias $\xi_0, \xi_1, \ldots, \xi_n$ con valores medios nulos y con *una matriz de correlación* $\{R_{kj}\}$ cuyas componentes son

$$R_{kj} = \mathbf{M}\xi_k \xi_j \quad (R_{kk} = \sigma_k^2;\ R_{kj} = \sigma_k r_{kj} \sigma_j)$$
$$(k, j = 0, 1, \ldots, n). \tag{4.28}$$

La matriz de correlación $\{R_{kj}\}$ está *determinada positivamente*: para cualesquiera valores reales de $c_0, c_1, \ldots, c_n$

$$\sum_{k,\,j=0}^{n} c_k c_j R_{kj} = \mathbf{M}\left(\sum_{k=0}^{n} c_k \xi_k\right)^2 \geqslant 0, \tag{4.29}$$

y con su ayuda, en el espacio lineal H de todas las magnitudes aleatorias de la forma

$$\xi = \sum_{k=0}^{n} c_k \xi_k \tag{4.30}$$

se puede introducir el producto escalar, determinándolo para cualesquiera $\xi' = \sum_{k=0}^{n} c'_k \xi_k$ y $\xi'' = \sum_{k=0}^{n} c''_k \xi_k$ como

$$(\xi', \xi'') = \mathbf{M}\xi'\xi'' = \sum_{k,\,j=0}^{n} c'_k c''_j R_{kj}. \tag{4.31}$$

La distancia media cuadrática, introducida anteriormente, es la distancia en el espacio de Euclides H con el producto escalar (4.31), a saber:

$$\|\xi' - \xi''\| = \sqrt{(\xi' - \xi'', \xi' - \xi'')}.$$

Como es conocido, en el espacio de Euclides H se tiene realmente a magnitud

$$\hat{\xi}_0 = \sum_{k=1}^{n} \hat{c}_k \xi_k,$$

que satisface la condición (4.27) planteada anteriormente,

$$\|\xi_0 - \hat{\xi}_0\| = \min_{c_1, \ldots, c_n} \left\|\xi_0 - \sum_{k=1}^{n} c_k \xi_k\right\|;$$

que representa en sí, geométricamente, una proyección del elemento $\xi_0 \in H$ sobre el subespacio de todos los elementos de la forma $\sum_{k=1}^{n} c_k \xi_k$.

Esta magnitud $\hat{\xi}_0$ puede ser hallada de las condiciones de ortogonalidad de la diferencia $\xi_0 - \hat{\xi}_0$ respecto a todas las magnitudes ξ_k, $k = 1, \ldots, n$, lo que da el siguiente sistema de ecuaciones lineales para hallar los coeficientes $\hat{c}_k$, $k = 1, \ldots, n$:

$$\left(\xi_0 - \sum_{k=1}^{n} \hat{c}_k \xi_k, \xi_j\right) = 0, \quad j = 1, \ldots, n,$$

o más detalladamente,

$$\sum_{k=1}^{n} R_{kj} \hat{c}_k = R_{0j}, \quad j = 1, \ldots, n. \tag{4.32}$$

(Ciertamente, desde el mismo comienzo, es natural examinar el caso de las magnitudes linealmente independientes $\xi_1, \ldots, \xi_n$ para las cuales el determinante del sistema (4.32) es distinto de cero.)

5. **Algunos teoremas sobre la convergencia.** Examinemos la sucesión de magnitudes aleatorias ξ_n, $n = 1, 2, \ldots$. Se dice que la magnitud aleatoria ξ es *el límite en la media cuadrática de la* sucesión ξ_n:

$$\text{l.i.m}_{n\to\infty} \xi_n = \xi,$$

si $\| \xi_n - \xi \| = \sqrt{\mathbf{M} | \xi_n - \xi |^2} \to 0$ para $n \to \infty$.

Para la sucesión ξ_n, $n = 1, 2, \ldots$ convergente en su media cuadrática, según la desigualdad de Chebishev también tenemos *la convergencia por la probabilidad*

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \| \xi_n - \xi \|^2 \to 0 \text{ para } n \to \infty,$$

es decir, por pequeños que fuesen $\varepsilon > 0$ y $\delta > 0$, para los valores de n suficientemente grandes, la probabilidad de cada uno de los sucesos separados $\{|\xi_n - \xi| > \varepsilon\}$ será menor que δ. De tal modo, si despreciamos el suceso poco probable $A_n = \{|\xi_n - \xi| > \varepsilon\}$, al examinar la magnitud aleatoria ξ_n, entonces tendremos que $\xi_n \approx \xi$ con una exactitud hasta de ε.

Recordemos que examinamos las magnitudes aleatorias $\xi_n = \xi_n(\omega)$, $n = 1, 2, \ldots$, dependientes del resultado elemental $\omega \in \Omega$. Para tal o cual resultado elemental ω tenemos la sucesión correspondiente de números $\xi_n(\omega)$, $n = 1, 2, \ldots$, llamada *sucesión seleccionada*.

Fijemos un determinado ε. Como ya fue señalado, para la sucesión ξ_n, $n = 1, 2, \ldots$, convergente en la media cuadrática, cada suceso por separado $A_n \subseteq \Omega$, para el cual $|\xi_n(\omega) - \xi(\omega)| > \varepsilon$, es poco probable para los grandes valores de n ($\mathbf{P}(A_n) \to 0$ para $n \to \infty$). Sin embargo, puede ocurrir que en la sucesión A_n, $n = 1, 2, \ldots$, se realicen un número infinito de sucesos, es decir, en la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \ldots$, habrá un número infinito de miembros (designémoslos por $\xi_{n_k}(\omega)$) tales que $|\xi_{n_k}(\omega) - \xi(\omega)| > \varepsilon$ y, por consiguiente, tal sucesión $\xi_n(\omega)$, $n = 1, 2, \ldots$, en general no converge hacia el valor $\xi(\omega)$ de la magnitud límite ξ. Más adelante, se da un ejemplo sencillo de una sucesión de magnitudes aleatorias ξ_n, $n = 1, 2, \ldots$, que converge en la media cuadrática, para las cuales la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \ldots$, no converge para ninguno de los resultados elementales ω.

Ejemplo. Supongamos que el espacio de los resultados elementales ω es el segmento $0 \leqslant \omega \leqslant 1$ con una distribución uniforme de probabilidades y la sucesión ξ_n, $n = 1, 2, \ldots$, está compuesta de magnitudes aleatorias $\xi_{mk} = \xi_{mk}(\omega)$ de la forma

$$\xi_{mk}(\omega) = \begin{cases} 1, \text{ si } \dfrac{k-1}{m} < \omega \leqslant \dfrac{k}{m} \\ 0, \text{ para los restantes } \omega, \end{cases}$$

donde $k = 1, \ldots, m$, $m = 1, 2, \ldots$. La sucesión de magnitudes aleatorias ξ_n, $n = 1, 2, \ldots$, converge hacia cero en la media cuadrática, ya que $\mathbf{M}\xi_{mk}^2 = \dfrac{1}{m} \to 0$ para $m \to \infty$. Al mismo tiempo, para cualquier resultado elemental ω se tiene un número infinito de valores $\xi_{mk}(\omega)$ con los números m, k: $\dfrac{k-1}{m} < \omega \leqslant \dfrac{k}{m}$ tales que $\xi_{mk} = 1$. Se ve que para ninguno de los resultados elementales ω, la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \ldots$, no converge hacia cero.

Se dice, que *la sucesión de magnitudes aleatorias* ξ_n, $n = 1, 2, \ldots$, *converge hacia la magnitud aleatoria* ξ *con una probabilidad* 1, si para la sucesión seleccionada $\xi_n(\omega)$, $n = 1, 2, \ldots$,

$$\lim_{n \to \infty} \xi_n(\omega) = \xi(\omega) \tag{4.33}$$

y para cualquier resultado elemental ω, a excepción, puede ser, de algunos resultados ω que tienen en su conjunto la probabilidad 0, es decir, si la relación límite (4.33) es un suceso de probabilidad 1.

La convergencia $\xi_n(\omega) \to \xi(\omega)$ significa, que $|\xi_n(\omega - \xi(\omega)| \leqslant \varepsilon$ con cualquier $\varepsilon > 0$ para todos los n suficientemente grandes, en otras palabras, $|\xi_n(\omega) - \xi(\omega)| > \varepsilon$ solamente para un número finito de miembros de la sucesión $\xi_n(\omega)$, $n = 1, 2, \ldots$. De tal modo, si $\xi_n(\omega) \to \xi(\omega)$ con la probabilidad 1, entonces *para cualquier*

$\varepsilon > 0$ *se realiza con la probabilidad* 1, *sólo un número finito de sucesos* $A_n^\varepsilon = \{|\xi_n(\omega) - \xi(\omega)| > \varepsilon\}$, $n = 1, 2, \ldots$

Mostremos que ésta es la condición suficiente para que $\xi_n(\omega) \to \xi(\omega)$ cuando $n \to \infty$ con la probabilidad 1. Supongamos que $\varepsilon = \frac{1}{k}$ y designemos por B_k el suceso que significa que en la sucesión A_n^ε, $n = 1, 2, \ldots$, se realiza sólo un número finito de sucesos $A_n^\varepsilon = \left\{|\xi_n(\omega) - \xi(\omega)| > \frac{1}{k}\right\}$, $n = 1, 2, \ldots$. Evidentemente, B_k, $k = 1, 2, \ldots$, es una sucesión monótona decreciente de sucesos $B_1 \supseteq B_2 \supseteq \ldots$, y para el suceso $B = \bigcap_k B_k$, se deduce de la propiedad de continuidad, que $\mathbf{P}(B) = \lim_{k\to\infty} \mathbf{P}(B_k)$; para $\mathbf{P}(B_k) = 1$ tenemos $\mathbf{P}(B) = 1$; está claro que $\xi_n(\omega) \to \xi(\omega)$ cuando $n \to \infty$ para cualquier resultado elemental $\omega \in B$.

Lema (Borel—Cantelli). *Sea* A_n, $n = 1, 2, \ldots$, *una sucesión de sucesos arbitrarios* A_n *de probabilidad* $\mathbf{P}(A_n) = p_n$. *Si*

$$\sum_{n=1}^{\infty} p_n < \infty, \tag{4.34}$$

entonces ocurre, con la probabilidad 1 *sólo un número finito de tales sucesos.*

Demostración: Supongamos que B significa que ocurre un número finito de sucesos, entre los A_n, $n = 1, 2, \ldots$, y B_n significa que ocurre aunque sea sólo un suceso A_k, $k \geqslant n$ $\left(B_n = \bigcup_{k\geqslant n} A_k\right)$. Está claro, que el suceso B aparece cuando, y sólo cuando ocurren todos los sucesos B_n, $n = 1, 2, \ldots$ $\left(B = \bigcap_n B_n\right)$. Ya que $B_1 \supseteq B_2 \supseteq \ldots$, entonces $\mathbf{P}(B) = \lim_{n\to\infty} \mathbf{P}(B_n)$. La probabilidad de la unión de cualesquiera sucesos siempre no supera la suma de sus probabilidades de modo que

$$\mathbf{P}(B_n) \leqslant \sum_{k=n}^{\infty} \mathbf{P}(A_k) = \sum_{k=n}^{\infty} p_k,$$

donde según la condición (4.34) $\sum_{k=n}^{\infty} p_k \to 0$ para $n \to \infty$. Por consiguiente,

$$\mathbf{P}(B) = \lim_{n\to\infty} \mathbf{P}(B_n) = 0$$

y el suceso $\overline{B}$ complementario del B, que significa que en la sucesión $A_1, A_2, \ldots$ ocurre sólo un número finito de sucesos, tiene una probabilidad $\mathbf{P}(\overline{B}) = 1$.

Teorema 1. *Si la sucesión de magnitudes aleatorias* ξ_n, $n = 1, 2, \ldots$, *converge en su media cuadrática hacia la magnitud aleatoria* ξ.

«bastante rápidamente», a saber, si

$$\sum_{k=1}^{\infty} \| \xi_n - \xi \|^2 < \infty,$$

entonces ella converge también con la probabilidad 1.

Demostración. Para cualquier $\varepsilon > 0$ en el suceso $A_n = \{| \xi_n(\omega) - \xi(\omega) | > \varepsilon\}$, tenemos según la desigualdad de Chebishev que $\mathbf{P}(A_n) \leqslant \| \xi_n - \xi \|^2/\varepsilon^2$ y, *por consiguiente*,

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) \leqslant \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \| \xi_n - \xi \|^2 < \infty.$$

Según el lema de Borel—Cantelli ocurre, con la probabilidad 1, solamente un número finito de sucesos A_n, $n = 1, 2, \ldots$, y esto significa, como indicamos anteriormente, que $\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)$ con la probabilidad 1.

La distancia media cuadrática $\| \xi_1 - \xi_2 \| = \sqrt{\mathbf{M} | \xi_1 - \xi_2 |^2}$, como corresponde a una verdadera distancia entre los puntos del espacio H, (véase el punto 4 anterior), satisface *la desigualdad del triángulo*: para tres puntos cualesquiera $\xi, \eta, \zeta \in H$, la distancia entre ξ, η no supera a la suma de las distancias entre ξ, ζ y η, ζ:

$$\| \xi - \eta \| \leqslant \| \xi - \zeta \| + \| \eta - \zeta \|. \tag{4.35}$$

Esto se deduce directamente de la desigualdad general (4.14) aplicada a la distancia $\xi_1 = \xi - \zeta$ y $\xi_2 = \eta - \zeta$.

Se cumple la siguiente proposición.

Teorema 2. *La sucesión de magnitudes aleatorias* ξ_n, $n = 1, 2, \ldots$, *de* H *converge en su media cuadrática hacia una magnitud aleatoria* $\xi \in H$ *cuando, y sólo cuando*

$$\lim_{n,\, m\to\infty} \| \xi_n - \xi_m \| = 0. \tag{4.36}$$

La necesidad de la condición (4.36) se deduce de la desigualdad del triángulo: si la sucesión ξ_n, $n = 1, 2, \ldots$, converge en la media cuadrática hacia la magnitud ξ, entonces

$$\| \xi_n - \xi_m \| \leqslant \| \xi_n - \xi \| + \| \xi_m - \xi \| \to 0 \text{ para } n, m \to \infty.$$

La demostración de que para la condición (4.36) existe la magnitud $\xi \in H$ tal que $\xi = \lim_{n\to\infty} \xi_n$, exige la aplicación de un aparato matemático especial (teoría de la medida), y por eso omitimos la demostración[1]).

[1]) Véase, por ejemplo, el libro de I. I. Guijman, A. V. Skorojod (pág. 108). Señalemos que el teorema 2, nosotros lo utilizamos, sólo al determinar las llamadas integrales estocásticas (véanse las págs. 204, 225, 234).

Examinemos la sucesión ξ_n, $n = 1, 2, \ldots$, que converge en la media cuadrática hacia la magnitud aleatoria ξ. Las magnitudes $\xi_n \in H$ tienen unas esperanzas matemáticas finitas $\mathbf{M}\xi_n$, ya que según la desigualdad (4.13) $\mathbf{M} \mid \xi_n \mid \leqslant \sqrt{\mathbf{M}\xi_n^2} < \infty$. Mostremos que

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi. \tag{4.37}$$

En efecto, como ya se indicó, la magnitud $\xi \in H$ tiene una esperanza matemática finita $\mathbf{M}\xi$ y

$$\mid \mathbf{M}\xi_n - \mathbf{M}\xi \mid \leqslant \mathbf{M} \mid \xi_n - \xi \mid \leqslant \sqrt{\mathbf{M} \mid \xi_n - \xi \mid^2} \to 0$$

para $n \to \infty$.

Sea η_n, $n = 1, 2, \ldots$, otra sucesión convergente en la media cuadrática (hacia la magnitud aleatoria η).

Teorema 3. *Tiene lugar la siguiente relación límite*

$$\lim_{n\to\infty} \mathbf{M}\xi_n\eta_n = \mathbf{M}\xi\eta. \tag{4.38}$$

Demostración. Para cualesquiera magnitudes $\xi, \eta \in H$ existe la esperanza matemática finita $\mathbf{M}\xi\eta$, ya que según la igualdad de Cauchy—Bunjakovsky $\mathbf{M} \mid \xi\eta \mid \leqslant \sqrt{\mathbf{M} \mid \xi \mid^2} \cdot \sqrt{\mathbf{M} \mid \eta \mid^2}$. Luego, los valores de las medias cuadráticas $\| \xi_n \| = \sqrt{\mathbf{M}\xi_n^2}$ y $\| \eta_n \| = \sqrt{\mathbf{M}\eta_n^2}$ están limitadas por una determinada constante C:

$$\| \xi_n \| \leqslant \| \xi \| + \| \xi_n - \xi \| \leqslant C$$
$$\| \eta_n \| \leqslant \| \eta \| + \| \eta_n - \eta \| \leqslant C,$$

por cuanto $\| \xi_n - \xi \| \to 0$, $\| \eta_n - \eta \| \to 0$ para $n \to \infty$. Utilizando de nuevo la desigualdad de Cauchy—Bunjakovsky obtenemos

$$\begin{aligned} \mid \mathbf{M}\xi_n\eta_n - \mathbf{M}\xi\eta \mid &\leqslant \mid \mathbf{M}\xi_n\eta_n - \mathbf{M}\xi_n\eta \mid + \mid \mathbf{M}\xi_n\eta - \mathbf{M}\xi\eta \mid \leqslant \\ &\leqslant \mathbf{M}(\mid \xi_n \mid \cdot \mid \eta_n - \eta \mid) + \mathbf{M}(\mid \eta \mid \cdot \mid \xi_n - \xi \mid) \leqslant \\ &\leqslant \| \xi_n \| \cdot \| \eta_n - \eta \| + \| \eta \| \cdot \| \xi_n - \xi \| \leqslant \\ &\leqslant C(\| \xi_n - \xi \| + \| \eta_n - \eta \|) \to \end{aligned}$$

para $n \to \infty$ lo que se exigía demostrar.

Exponemos sin demostración[1]) dos teoremas más sobre la convergencia de las esperanzas matemáticas.

Sea $\xi_1 \leqslant \xi_2 \leqslant \ldots$ una sucesión monótona creciente de magnitudes aleatorias no negativas. Para cada resultado elemental ω

[1]) Véase el libro citado anteriormente de I. I. Guijman, A. V. Skorojod (pág. 102). El teorema 4 siguiente (más exacto, su corolario (4.39), sólo se utiliza para la demostración en las págs. 191 y 234 y el teorema 5 en las págs. 151 y 166.

la sucesión seleccionada monótona creciente $\xi_n(\omega)$, $n = 1, 2, \ldots$, tiene un límite finito o infinito $\xi(\omega) \geqslant 0$. Supongamos que las magnitudes ξ_n tienen una esperanza matemática finita $\mathbf{M}\xi_n$, $n = 1, 2, \ldots$. Para esta sucesión monótona creciente existe un límite finito o infinito $\lim\limits_{n\to\infty} \mathbf{M}\xi_n$.

Teorema 4. *Si la sucesión* $\mathbf{M}\xi_n$, $n = 1, 2, \ldots$, *es limitada, entonces la magnitud aleatoria límite* $\xi = \lim\limits_{n\to\infty} \xi_n$ *es finita con una probabilidad* 1 *y*

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi.$$

Como consecuencia de este teorema obtenemos el siguiente resultado.

Supongamos que la sucesión de magnitudes aleatorias no negativas ξ_k, $k = 1, 2, \ldots$, *sea tal que la serie de sus esperanzas matemáticas* $\sum\limits_{k=1}^{\infty} \mathbf{M}\xi_k$ *es convergente. Entonces con la probabilidad* 1

$$\sum_{k=1}^{\infty} \xi_k < \infty$$

y

$$\mathbf{M}\left(\sum_{k=1}^{\infty} \xi_k\right) = \sum_{k=1}^{\infty} \mathbf{M}\xi_k. \tag{4.39}$$

En efecto, las sumas parciales $\sum\limits_{k=1}^{n} \xi_k$, $n = 1, 2, \ldots$, forman una sucesión monótona creciente, para la cual las esperanzas matemáticas $\mathbf{M}\left(\sum\limits_{k=1}^{n} \xi_k\right) = \sum\limits_{k=1}^{n} \mathbf{M}\xi_k$, $n = 1, 2, \ldots$, están limitadas: $\mathbf{M}\left(\sum\limits_{k=1}^{n} \xi_k\right) \leqslant \sum\limits_{k=1}^{\infty} \mathbf{M}\xi_k < \infty$, y según el teorema 4 existe, con la probabilidad 1, el límite finito $\lim\limits_{n\to\infty} \sum\limits_{k=1}^{n} \xi_k = \sum\limits_{k=1}^{\infty} \xi_k$ siendo

$$\mathbf{M}\left[\lim_{n\to\infty} \sum_{k=1}^{n} \xi_k\right] = \lim_{n\to\infty} \mathbf{M}\left[\sum_{k=1}^{n} \xi_k\right] = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{M}\xi_k.$$

Teorema 5. *Supongamos que la sucesión de magnitudes aleatorias* ξ_n, $n = 1, 2, \ldots$, *converge, con la probabilidad* 1, *hacia una mag-*

nitud ξ, *siendo* $|\xi_n| \leqslant \eta$, *donde la magnitud aleatoria* η *tiene una esperanza matemática finita* ($\mathbf{M}|\eta| < \infty$). *Entonces las magnitudes* ξ_n *y* ξ *también tienen esperanzas matemáticas, siendo*

$$\lim_{n\to\infty} \mathbf{M}\xi_n = \mathbf{M}\xi.$$

§ 5. SERIES INFINITAS DE PRUEBAS INDEPENDIENTES Y LEY DE LOS GRANDES NUMEROS

1. Ley de los grandes números. Examinemos la sucesión de magnitudes aleatorias independientes

$$\xi_1, \xi_2, \ldots, \tag{5.1}$$

que tienen una misma distribución de probabilidades (para claridad nos podemos imaginar que se ejecuta una serie de pruebas iguales independientes $\Omega_1, \Omega_2, \ldots$, y en cada prueba Ω_n se observa la magnitud correspondiente ξ_n, $n = 1, 2, \ldots$).

Supongamos que existe la esperanza matemática y la dispersión:

$$a = \mathbf{M}\xi_n, \quad \sigma^2 = \mathbf{D}\xi_n, \quad n = 1, 2, \ldots,$$

y examinemos el valor medio «empírico»

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k = \frac{S_n}{n} \tag{5.2}$$

(que el «observador» puede calcular después de realizarse las pruebas $\Omega_1, \ldots, \Omega_n$). Para la suma $S_n = \sum_{k=1}^{n} \xi_k$ tenemos

$$\mathbf{M}\frac{S_n}{n} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{M}\xi_k = a,$$

$$\mathbf{D}\frac{S_n}{n} = \mathbf{M}\left(\frac{S_n}{n} - a\right)^2 = \frac{1}{n^2} = \sum_{k=1}^{n} \mathbf{D}\xi_k = \frac{\sigma^2}{n}.$$

Vemos que

$$\left\|\frac{1}{n}\sum_{k=1}^{n} \xi_k - a\right\| = \frac{\sigma}{\sqrt{n}} \to 0 \text{ para } n \to \infty, \tag{5.3}$$

es decir, *la media empírica* $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ *converge en la media cuadrática hacia la esperanza matemática* a:

$$\text{l.i.m}_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\xi_k = a.$$

De tal modo, *para cualquier* $\varepsilon > 0$ *la probabilidad de que la media empírica* $\frac{1}{n}\sum_{k=1}^{n}\xi_k$, *para los valores* n *suficientemente grandes, se diferencia de la esperanza matemática* a *en no más de* ε, *es casi igual a* 1, *más exacto,*

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| \leqslant \varepsilon\right\} \geqslant 1-\delta$$

siendo $\delta > 0$ *tan pequeño como se quiera, para los valores* n *suficientemente grandes; con ello la desigualdad de Chebishev da la siguiente valuación de la probabilidad indicada:*

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| \leqslant \varepsilon\right\} \geqslant 1-\frac{\sigma^2}{n\varepsilon^2}. \tag{5.4}$$

En la vida real siempre despreciamos la posibilidad de que aparezcan los sucesos, cuya probabilidad es suficientemente pequeña. Si despreciamos el suceso poco probable $\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| > \varepsilon$ (la probabilidad no es mayor de un δ suficientemente pequeño), entonces se puede considerar, que para cualquier n fijado suficientemente grande digamos $n \geqslant \frac{\sigma^2}{\delta\varepsilon^2}$

$$a-\varepsilon \leqslant \frac{1}{n}\sum_{k=1}^{n}\xi_k \leqslant a+\varepsilon,$$

de modo que, con una exactitud hasta de ε, la media empírica $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ coincide con el valor medio de a:

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \approx a. \tag{5.5}$$

Esta ley interesante (expresada por las relaciones (5.3) y (5.4) lleva el nombre de *ley de los grandes números.*

Propiamente hablando, la ley indicada de los grandes números la hemos deducido, no sólo para las magnitudes aleatorias independientes, sino también para las magnitudes aleatorias no correlativas ξ_k, $k = 1, 2, \ldots$ ($\mathbf{M}\xi_k = a$, $\mathbf{D}\xi_k = \sigma^2$). Para tales magnitudes también se puede obtener el siguiente resultado:

Teorema (ley reforzada de los grandes números). *Con la probabilidad* 1

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \to a \text{ para } n \to \infty. \tag{5.6}$$

Demostración[1]). Sin limitar la generalidad se puede considerar $a = 0$. Según la desigualdad de Chebishev para cualquier $\varepsilon > 0$

$$\sum_{m=1}^{\infty}\mathbf{P}\left\{\left|\frac{S_{m^2}}{m^2}\right| > \varepsilon\right\} \leqslant \frac{m^2\sigma^2}{m^4\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}\cdot\frac{1}{m^2},$$

de modo que

$$\sum_{m=1}^{\infty}\mathbf{P}\left\{\left|\frac{S_{m^2}}{m^2}\right| > \varepsilon\right\} \leqslant \frac{\sigma^2}{\varepsilon^2}\sum_{m=1}^{\infty}\frac{1}{m^2} < \infty;$$

según el lema de Borel—Cantelli, con la probabilidad 1 puede ocurrir un solo número finito de sucesos $A_m = \left\{\left|\frac{S_{m^2}}{m^2}\right| > \varepsilon\right\}$, $m = 1, 2, \ldots$ y de tal modo $\frac{S_{m^2}}{m^2} \to 0$ con la probabilidad 1 (para $m \to \infty$). Después, sea

$$\eta_m = \max_{m^2+1 \leqslant k \leqslant (m+1)^2} |\xi_{m^2} + \ldots + \xi_k|.$$

Evidentemente,

$$\mathbf{P}\left\{\left|\frac{\eta_m}{m^2}\right| > \varepsilon\right\} \leqslant \sum_{k=m^2+1}^{(m+1)^2}\mathbf{P}\left\{\left|\frac{\xi_{m^2}+\ldots+\xi_k}{m^2}\right| > \varepsilon\right\} \leqslant$$

$$\leqslant \sum_{k=m^2+1}^{(m+1)^2}\frac{(k-m^2)\sigma^2}{m^4\varepsilon^2} \leqslant 2m\cdot\frac{2m\sigma^2}{m^4\varepsilon^2} = \frac{4\sigma^2}{\varepsilon^2}\cdot\frac{1}{m^2}$$

y

$$\sum_{m=1}^{\infty}\mathbf{P}\left\{\left|\frac{\eta_m}{m^2}\right| > \varepsilon\right\} \leqslant \frac{4\sigma^2}{\varepsilon^2}\sum_{m=1}^{\infty}\frac{1}{m^2} < \infty;$$

[1]) La demostración expuesta aquí ha sido comunicada al autor por Yu. V. Prójorov.

por consiguiente, con la probabilidad 1, puede ocurrir, sólo un número finito de sucesos $B_m = \left\{ \left| \frac{\eta_m}{m^2} \right| > \varepsilon \right\}$, $m = 1, 2, \ldots$, y de tal modo, $\frac{\eta_m}{m^2} \to 0$ con la probabilidad 1 (para $m \to \infty$). Como resumen, para cualquier n, $m^2 \leqslant n \leqslant (m+1)^2$, tenemos

$$\left| \frac{S_n}{n} \right| \leqslant \left| \frac{S_{m^2}}{m^2} \right| + \frac{\eta_m}{m^2} \to 0$$

con la probabilidad 1 para $n \to \infty$, lo que se exigía demostrar.

También tiene lugar el siguiente resultado fundamental[1]).

Ley reforzada de los grandes números. Para la sucesión de magnitudes aleatorias independientes, igualmente distribuidas, $\xi_1, \xi_2, \ldots$, *que tienen una esperanza matemática finita* a, *con la probabilidad* 1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \xi_k = a. \tag{5.7}$$

2. Probabilidad y frecuencia. Dirijámonos a un determinado suceso A, ligado con una prueba determinada. Supongamos que se tiene la posibilidad de reproducir repetidas veces esta prueba, de modo que las pruebas Ω_1, Ω_2 son i n d e p e n d i e n t e s.

Introduzcamos las magnitudes aleatorias $\xi_1, \xi_2 \ldots$ donde cada magnitud ξ_k está solamente ligada con su prueba correspondiente Ω_k definida del siguiente modo: $\xi_k = 1$, si en la k-ésima prueba ocurre el suceso A, y $\xi_k = 0$ en caso contrario. Ahora tenemos

$$\mathbf{M}\xi_k = p, \quad \mathbf{D}\xi_k = p\,(1-p),$$

donde $p = \mathbf{P}(A)$ es la probabilidad del suceso A en cada prueba por separado.

La media empírica

$$\frac{1}{n} \sum_{k=1}^{n} \xi_n = \frac{n}{A}$$

representa *la frecuencia del suceso* A en las pruebas examinadas, que se define como la relación entre el número $n(A)$ de aquellas pruebas, en las cuales ocurrió este suceso y el número total n de pruebas. Según la ley de los grandes números

$$\frac{n(A)}{n} \to \mathbf{P}(A) \text{ para } n \to \infty \tag{5.8}$$

[1]) Véase, por ejemplo, el libro de V. Feller «Introducción a la teoría de probabilidades y sus aplicaciones», tomo I, II, Moscú 1964 (pág. 266). Aquí se tiene en cuenta que las magnitudes $\xi_1, \xi_2, \ldots$, no tienen limitada obligatoriamente la dispersión σ^2.

de manera que en *una gran serie de pruebas independientes, la frecuencia* $\frac{n(A)}{n}$ *del suceso A coincide, prácticamente, con la probabilidad de este suceso:*

$$\frac{n(A)}{n} \approx \mathbf{P}(A). \tag{5.9}$$

Esta ley (que recibió su afirmación en los múltiples experimentos de la práctica) nos da una base para valorar experimentalmente, en las aplicaciones la probabilidad conocida $\mathbf{P}(A)$ de tal o cual suceso A, como $\mathbf{P}(A) \approx \frac{n(A)}{n}$ (véase el ejemplo en la pág. 23).

Más adelante se expone la tabla 1 (tomada del libro, citado anteriormente, de W. Feller, véase la pág. 33) en donde se dan los resultados de 100 series con 100 pruebas, en cada una de las cuales se lanzó una moneda simétrica. De esta tabla se ve que la frecuencia $\frac{n(A)}{n}$ de la caída con «cruz» (suceso A) en cada serie coincide con gran exactitud con la probabilidad $\mathbf{P}(A) = 1/2$.

Tabla 1

Número de pruebas	Número de cruces										Número total de cruces
0—1000	54	46	53	55	46	54	41	48	51	53	501
2000	48	46	40	53	49	49	48	54	53	45	485
3000	43	52	58	51	51	50	52	50	53	49	509
4000	58	60	54	55	50	48	47	57	52	55	536
5000	48	51	51	49	44	52	50	46	53	41	485
6000	49	50	45	52	52	48	47	47	47	51	488
7000	45	47	41	51	49	59	50	55	53	50	500
8000	53	52	46	52	44	51	48	51	46	54	497
9000	45	47	46	52	47	48	59	57	45	48	494
10 000	47	41	51	48	59	51	52	55	39	41	484

CAPITULO II

ALGUNAS DISTRIBUCIONES DE PROBABILIDADES

§ 1. ELECCION ALEATORIA Y VARIACION ALEATORIA

1. Algunas fórmulas combinatorias. Al calcular las probabilidades aportan una gran ayuda las fórmulas combinatorias. Expongamos las más importantes de ellas.

Combinaciones de elementos elegidos de diferentes grupos. Supongamos que se tiene r grupos diferentes, compuestos por determinados elementos. El primer grupo contiene n_1 elementos $a_1, \ldots, a_{n_1}$; el segundo contiene n_2 elementos $b_1, b_2, \ldots, b_{n_2}$; $\ldots$ el último grupo r contiene n_r elementos $c_1, c_2, \ldots, c_{n_r}$. Se componen todas las combinaciones posibles de r elementos pertenecientes a los diferentes grupos de tal modo que en una combinación separada entra sólo un elemento de cada grupo. Ellas tienen la forma

$$(a, b, \ldots, c).$$

Las combinaciones $(a, b, \ldots, c)$ y $(\tilde{a}, \tilde{b}, \ldots, \tilde{c})$ se consideran distintas, si se tiene aunque un solo par de elementos distintos entre sí a y $\tilde{a}$, b y $\tilde{b}$, $\ldots$, c y $\tilde{c}$. El número de tales combinaciones es

$$N = n_1 \cdot n_2, \ldots \cdot n_r \qquad (1.1)$$

(para $r = 2$ la fórmula (1.1) es evidente, y para un número arbitrario r se establece fácilmente por inducción).

Elección con retorno. Supongamos que se tiene n objetos diferentes $a_1, a_2, \ldots, a_n$. Elijamos sucesivamente r objetos de este conjunto de tal modo que cada objeto elegido se fija y se retorna atrás. Resultado de tal elección es la combinación de la forma

$$(a_{i_1}, a_{i_2} \ldots, a_{i_r}).$$

Las combinaciones $(a_{i_1}, a_{i_2}, \ldots, a_{i_r})$ y $(a_{j_1}, a_{j_2}, \ldots, a_{j_r})$ se consideran distintas, si en cualquier paso fueron elegidos distintos objetos, es decir $a_{i_k} \neq a_{j_k}$, aunque sea para un solo k. Nos podemos imaginar que se tienen r grupos iguales de n elementos, y en el paso k se elige

el elemento del k-ésimo grupo. En este caso particular, la fórmula (1.1) da la siguiente expresión para el número de combinaciones diferentes N:

$$N = n^r. \tag{1.2}$$

Número de variaciones. Elección sin retorno. Supongamos que r objetos diferentes se colocan por determinadas celdas. Con ello las celdas son distinguibles por el observador y su número es igual a n. Supongamos que los objetos se colocan de tal modo que en cada celda cae no más de un objeto. Numeremos todos los objetos que se tienen y las celdas. Entonces cada variación la podemos describir por la combinación $(i_1, i_2, \ldots, i_r)$, donde i_1 es el número de la celda en donde cae el 1er objeto, i_2 es el número de la celda donde cae el 2° objeto, etc., i_r es el número de la celda donde cae el último objeto r. En la combinación $(i_1, i_2, \ldots i_r)$ el elemento i_1 puede ser cualquiera de los n elementos (celdas), i_2 es uno de los $n - 1$ elementos restantes, etc., i_r puede ser uno de los $n - r + 1$ elementos posibles. De acuerdo con la fórmula general (1.1) en total existen

$$N = n\,(n - 1)\,\ldots\,(n - r + 1) \tag{1.3}$$

distintas variaciones de r elementos por n celdas.

Supongamos, que se tiene el conjunto de n elementos diferentes $a_1, a_2, \ldots, a_n$. De este conjunto se eligen r elementos. Se puede considerar que ellos se eligen a la vez y se colocan en un orden determinado, o se puede considerar que los elementos se eligen sucesivamente uno después de otro, con ello el elemento elegido no se retorna atrás. Como resultado de tal elección se forma la combinación $(a_{i_1}, a_{i_2}, \ldots$ $\ldots, a_{i_r})$, donde a_{i_1} es el primero de los elementos elegidos, a_{i_2} es el segundo de los elementos elegidos, etc., a_{i_r} es el último de los elementos elegidos; a_{i_1} puede ser cualquiera de los n elementos que se tienen, a_{i_2}, cualquiera de los $n - 1$ elementos restantes, etc., a_{i_r}, cualquiera de los $n - r + 1$ elementos posibles. En total existen

$$N = n\,(n - 1) \qquad (n - r + 1)$$

distintas elecciones $(a_{i_1}, a_{i_2}, \ldots, a_{i_r})$ de r objetos, elegidos, sin retorno, del conjunto de volumen n.

Para $r = n$, el número de variaciones coincide con *el número de permutaciones de n elementos*:

$$N = n\,(n - 1) \quad \ldots 1 = n! \tag{1.4}$$

Número de combinaciones. Supongamos que se colocan r objetos no distinguibles entre sí por n celdas (en cada celda cae no más de un objeto). Entonces el número de variaciones distinguibles coincide con el número de grupos distintos por r celdas, que se puede formar del conjunto general de volumen n (en un grupo se agrupan todas las celdas ocupadas), y es igual *al número de combinaciones* C_n^r

de n elementos por r:

$$N = C_n^r = \frac{n!}{r!\,(n-r)!}\,. \tag{1.5}$$

Más general es el caso, cuando el conjunto de n elementos se divide en un número fijo k de grupos, siendo también fijo el número de elementos en cada grupo, digamos, en el grupo con el número i se tienen, justamente, n_i elementos. En lo restante, los elementos se agrupan lo más arbitrariamente. El número de combinaciones diferentes de n elementos por $n_1, n_2, \ldots, n_k$ elementos ($n_1 + n_2 + \ldots + n_k = n$) es

$$N = \frac{n!}{n_1!\,n_2!\ldots n_k!}\,. \tag{1.6}$$

La fórmula (1.5) puede ser obtenida de la siguiente forma. Supongamos temporalmente, que todos los objetos examinados son diferentes; entonces el número de variaciones diferentes (véase (1.3)) será igual a $n\,(n-1)\ldots(n-r+1)$. Para cualquier grupo $i_1, \ldots, i_r$ de celdas ocupadas, entre aquel número figuran también las variaciones que se obtienen permutando r objetos por las celdas indicadas; el total de tales permutaciones será $r!$. Por consiguiente, el número de combinaciones C_n^r de n celdas por r[1]), que nos interesa, es $r!$ veces menor que el número indicado de variaciones $n\,(n-1)\ldots\ldots(n-r+1)$, a saber:

$$C_n^r = \frac{n\,(n-1)\ldots(n-r+1)}{r!} = \frac{n!}{r!\,(n-r)!}\,.$$

Para la demostración de la fórmula (1.6) la volvemos a escribir en la forma:

$$N = C_n^{n_1} C_{n-n_1}^{n_2} \ldots C_{n-n_1-\ldots-n_{k-2}}^{n_{k-1}}.$$

La formación de grupos de $n_1, n_2, \ldots, n_k$ elementos se puede realizar consecutivamente. Primeramente, formemos el grupo de n_1 elementos (esto se puede hacer de $N_1 = C_n^{n_1}$ modos diferentes); luego de los $n - n_1$ elementos restantes formamos el grupo de n_2 elementos (esto se puede hacer de $N_2 = C_{n-n_1}^{n_2}$ modos diferentes), etc.; los elementos restantes al final $n - n_1 - \ldots - n_{k-2} = n_{k-1} + n_k$ se pueden dividir en dos grupos de n_{k-1} y n_k elementos, de $N_{k-1} = C_{n-n_1-\ldots-n_{k-2}}^{n_{k-1}}$ modos diferentes. Combinando a cada paso los modos de formación de grupos, según la fórmula general (1.1), en total obtenemos el número indicado $N = N_1 N_2 \ldots N_{k-1}$ de todas las combinaciones posibles.

[1]) El término C_n^r se suele leer así: «el número de combinaciones de n elementos (celdas) tomados de r en r». (*N. del T.*)

En todas las igualdades expuestas anteriormente, se encuentra la expresión $n! = 1 \cdot 2 \ldots (n-1)\, n$, para la cual se tiene *la siguiente fórmula de Stirling*[1]):

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}, \tag{1.7}$$

donde $e = 2{,}718\ldots$ es la base de los logaritmos naturales (aquí y en adelante la relación $\alpha_n \sim \beta_n$ entre las magnitudes variables α_n y β_n significa que $\frac{\alpha_n}{\beta_n} \to 1$ para $n \to \infty$). Más exacta que (1.7) es la relación

$$e^{\frac{1}{12n+1}} \leqslant \frac{n!}{\sqrt{2\pi n}\, n^n e^{-n}} \leqslant e^{\frac{1}{12n}},$$

justa para todos los $n = 1, 2, \ldots$

Ejemplo. Un grupo de pasajeros de r personas se sientan en el tren eléctrico suburbano, compuesto de $n \geqslant r$ vagones. Supongamos que cada pasajero elige su vagón casualmente y se encuentra con una misma probabilidad en cualquiera de los vagones. ¿Cuál será la probabilidad de que todos ellos entren en un mismo vagón? ¿Qué probabilidad de que entren en diferentes vagones?

Cada pasajero puede elegir uno de los n vagones, de modo que todas las posibles combinaciones $(i_1, i_2, \ldots, i_r)$, donde i_k es el número del vagón, elegido por el k-ésimo pasajero, alcanzan el valor $N = n^r$.

El suceso A, o sea «todos los r pasajeros entran en un vagón» ocurre en cualquiera de los $N(A) = n$ resultados, en los cuales $i_1 = i_2 = \ldots = i_r$, y su probabilidad es

$$\mathbf{P}(A) = \frac{N(A)}{N} = \frac{1}{n^{r-1}}.$$

El suceso B, o sea «en ninguno de los vagones entra más de uno de los pasajeros examinados», aparece cuando, y sólo cuando todos los elementos i_k, $k = 1, \ldots, r$, son distintos entre sí. Se puede considerar que se realiza la elección (s i n r e t o r n o) de r elementos del conjunto de volumen n: el primer pasajero elige cualquiera de los n vagones, el segundo, cualquiera de los $(n-1)$ vagones restantes, etc. Por consiguiente, el número de combinaciones $(i_1, i_2, \ldots, i_r)$ que conducen a la aparición del suceso B, es igual a $n(n-1)\ldots(n-r+1)$, y la probabilidad de este suceso es

$$\mathbf{P}(B) = \frac{n(n-1)\ldots(n-r+1)}{n^r}.$$

Ejemplo. Imaginémonos que se tienen n llaves, una de las cuales entra en la cerradura. Las llaves se prueban sucesivamente. ¿Cuál será la probabilidad de que la llave útil caiga en el paso k?

[1]) Véase, por ejemplo, el libro de V. Feller citado anteriormente.

Consideramos que las llaves están dispuestas en orden aleatorio $(i_1, i_2, \ldots, i_n)$ siendo todas estas disposiciones equiprobables. Supongamos que la llave útil tiene el número i. Entonces la probabilidad que nos interesa será igual a la probabilidad de que, en la disposición aleatoria de los números $1, \ldots, n$, esté situado el número i en el lugar k. El número de todas las variaciones posibles $(i_1, \ldots, i_n)$ que se tienen es igual a $n!$, y el número de variaciones con el elemento fijado $i_k = i$ es igual a $(n-1)!$, de modo que la probabilidad buscada es

$$\frac{(n-1)!}{n!} = \frac{1}{n}.$$

Ejemplo. Examinemos el juego de «préférence»[1]), cuando los 32 naipes mayores de la baraja de naipes están distribuidos de un modo aleatorio entre los tres jugadores, que recibieron 10 naipes, y «la toma» de dos cartas, que se le concede a uno de los jugadores que luego anuncia la categoría del juego. Supongamos que junto con «la toma» el jugador tiene 5 naipes mayores de un mismo palo (digamos copas), excluyendo «la dama». Al anunciar la categoría de juego hay que tener en cuenta la posibilidad de que alguno de los compañeros tenga los tres naipes restantes del palo de copas, incluyendo «la dama» (más abreviado, que tenga la «tercera dama»). ¿Cuál es la probabilidad de este suceso?

En total existen $C_{20}^{10} = \frac{20!}{10!\,10!}$ distribuciones equiprobables de 20 naipes en dos grupos de 10 naipes, que se entregan a cada dos compañeros jugadores. Si fijamos la combinación de «la tercera dama» en cualquier jugador determinado, entonces el número de distribuciones de naipes, compatibles con ésta, entre dos jugadores es igual al número de combinaciones $C_{17}^{7} = \frac{17!}{7!\,10!}$ de los 17 naipes restantes en grupos de 7 (o lo que es igual, de 17 tomados de 10 en 10). Por lo tanto, «la tercera dama» la tendrá un determinado jugador con la probabilidad

$$\frac{C_{17}^{7}}{C_{20}^{10}} = \frac{8 \cdot 9 \cdot 10}{18 \cdot 19 \cdot 20} = \frac{2}{19},$$

y cada uno de los dos jugadores, con la probabilidad doble, igual a $\frac{4}{19}$.

2. Distintas distribuciones de las partículas independientes en el espacio fásico. Examinemos el siguiente modelo sencillo. Imaginémonos el proceso del cocido de la masa para los bollitos. En la masa se echan las uvas pasas, luego todo se mezcla minuciosamente y como resultado de lo cual las uvas pasas que son «partículas» se distribuyen de un modo aleatorio dentro del «espacio fásico» ocupado por la masa.

[1]) El juego de «préférence» con naipes franceses es muy parecido al juego español del «tresillo». (*N. del T.*)

Supongamos que nuestro «espacio fásico» de volumen V (masa) se divide en n celdas iguales, lo suficientemente grandes en comparación con el volumen total v de las partículas (uvas pasas) existentes:

$$\frac{V}{n} \gg v.$$

¿Cómo se distribuyen aquí las partículas?

Parece ser bastante semejante a la verdad el hecho de que después de la mezcla repetida de la masa, todas las distribuciones posibles de las uvas pasas por cada una de las celdas es equiprobable. Supondremos que esto es realmente así. Si el número total de uvas pasas es r, entonces cada distribución se puede describir por la combinación $(i_1, \ldots, i_r)$, donde i_1 es el número de la celda, donde cae la primera uva pasa, i_2, el número de celda donde cae la segunda uva pasa, etc. Cada uno de los índices r $i_1, \ldots, i_r$ puede tener n valores distintos, de modo que el número total de combinaciones diferentes $(i_1, \ldots, i_r)$ es $N = n^r$, por consiguiente, cada distribución de partículas $(i_1, \ldots, i_r)$ tiene la probabilidad $1/n^r$. La probabilidad de que caigan r_1 uvas pasas en la primera celda, r_2 uvas pasas en la segunda $\ldots$, r_n uvas pasas en la celda n $(r_1 + r_2 \ldots + r_n = r)$, es

$$\frac{r!}{r_1!\, r_2! \ldots r_n!}\, n^{-r}$$

ya que $\frac{r!}{r_1!\, r_2! \ldots r_n!}$ es igual al número de todas las combinaciones posibles de n grupos de $r_1, r_2, \ldots, r_n$ elementos (véase la fórmula (I.6)).

Observemos ahora, que las uvas pasas son para el observador las partículas indistinguibles, y desde este punto de vista todas aquellas distribuciones $(i_1, i_2, \ldots, i_r)$ en las que caen r_1 partículas en la primera celda, r_2 partículas en la segunda $\ldots$, r partículas en la celda n, son i n d i s t i n g u i b l e s. Desde el comienzo se podía haber hablado sobre las distribuciones, cada una de las cuales se describe por el número r_i de partículas que cayeron en la celda i correspondiente. Si consideramos que todas aquellas distribuciones $(r_1, \ldots, r_n)$ son equiprobables, entonces cada resultado $(r_1, \ldots, r_n)$ tendrá, no la probabilidad indicada anteriormente $\frac{r!}{r_1! \ldots r_n!}\, n^{-r}$, sino otra distinta. Precisamente, el número de distribuciones distintas, cada una de las cuales se representa en adelante, esquemáticamente, como la variación de los asteriscos entre los palotes:

$$\underbrace{|* \ldots *|}_{r_1}\underbrace{|* \ldots *|}_{r_2} \ldots \underbrace{|* \ldots *|}_{r_n},$$

(1 sobre el primer palote, 2 sobre el segundo, n sobre el último)

coincide con el número de modos, en que se pueden colocar $n - 1$ palotes, (que describen a n celdas) y r asteriscos (uvas pasas). Este

número, evidentemente, es igual al número de combinaciones $C_{n+r-1}^{n-1} = \frac{(n+r-1)!}{(n-1)!\,r!}$ de $n-1+r$ elementos en grupos de $n-1$ y, por consiguiente, cada resultado por separado $(r_1, \ldots, r_n)$ tendrá la probabilidad

$$\frac{(n-1)!\,r!}{(n+r-1)!}.$$

¿Cuál de las dos soluciones dadas antes es verdadera? Si consideramos que el «mecanismo aleatorio», descrito, anteriormente, (del entremezclado repetido), actúa de tal modo que cada uva pasa, por separado, cae en cualquiera de las n celdas con la misma probabilidad, siendo prácticamente, independiente del comportamiento de las otras uvas pasas, entonces la probabilidad de la distribución $(r_1, \ldots, r_n)$ será

$$P(r_1, \ldots, r_n) = \frac{r!}{r_1!\,r_2! \ldots r_n!}\, n^{-r}.$$

Esto se ve especialmente claro en el caso, cuando $n = r = 2$. Cada una de las dos uvas pasas (independiente una de otra) cae en dos cualesquiera celdas con la probabilidad 1/2 y

$$P(2, 0) = P(0, 2) = 1/4,$$
$$P(1, 1) = 1/2,$$

de modo que las posibles distribuciones (2, 0), (0, 2), (1, 1) no son de ningún modo equiprobables.

En el caso de distribución de las uvas pasas en la masa, el modelo de distribuciones equiprobables $(i_1, \ldots, i_r)$ descrito antes, evidentemente, no es satisfactorio, si el volumen de cada celda por separado es menor que el volumen de todas las partículas, o sea de las uvas pasas: (por ejemplo, en una celda no caben todas las partículas). Examinemos un nuevo modelo, en el cual las celdas elegidas son tan pequeñas (comparables con las medidas de las mismas uvas pasas) que en cada celda no puede caer más de una partícula.

Para la limitación indicada, cada distribución de las partículas puede ser descrita por la combinación $(i_1, \ldots, i_r)$ indicando las celdas ocupadas $i_1, \ldots, i_r$ (se considera naturalmente, que $r \leqslant n$) y el número total de distribuciones distintas es igual al número de combinaciones C_n^r con n elementos en grupos de r. Si consideramos que todas las distribuciones posibles son equiprobables (así está planteada la cuestión en nuestro modelo con las uvas pasas), entonces la probabilidad de cada distribución por separado $(i_1, \ldots, i_r)$ es

$$\frac{(n-r)!\,r!}{n!}.$$

La cuestión sobre la distribución de las partículas independientes en el espacio fásico (uvas pasas en la masa) puede ser planteada del

siguiente modo. Se elige una zona determinada del espacio fásico de volumen v. ¿Cuál será la probabilidad de que caiga en esta zona tal o cual número de partículas?

Dividamos imaginariamente todo el espacio fásico en celdas tan pequeñas que en cada una de ellas puede caer no más de una partícula. Consideremos igual que anteriormente, que todas las posibles distribuciones de partículas $(i_1, i_2, \ldots, i_r)$, donde i_r es el número de la celda ocupada por la partícula r, son equiprobables. Si consideramos, convencionalmente, las celdas ocupadas y libres como bolas blancas y negras, entonces nos podemos imaginar al espacio fásico como una determinada caja, que contiene r bolas blancas y $n - r$ negras, que se han remezclado repetida y minuciosamente (r es el número de todas las partículas en el espacio fásico, n es el número de todas las celdas, $n \geqslant r$).

Sea m el número de celdas de la zona designada en el espacio fásico v y ξ es el número de partículas que caen, según el caso, en esta zona (número de celdas ocupadas entre las m indicadas). La probabilidad $P_\xi(k)$ de que el número ξ sea igual a k, $0 \leqslant k \leqslant m$, es la misma que para el número ξ de bolas blancas, que suelen encontrarse entre las m bolas sacadas al azar durante su elección aleatoria de la caja (urna) con r bolas blancas y $n - r$ bolas negras.

La expresión «sacarla al azar» significa más exactamente, que m bolas cualesquiera se eligen con la misma probabilidad (si se numeran todas las bolas existentes, se eligen con la misma probabilidad cualesquiera bolas $j_1, \ldots, j_m$). El número N de resultados distintos (cada uno de los cuales representa en sí a la elección de tales o cuales bolas $j_1, \ldots, j_m$) es igual al número de combinaciones $C_n^m = \frac{n!}{m!\,(n-m)!}$. A nosotros nos interesa el suceso A, que consiste en que entre las m bolas sacadas se encuentren justamente k bolas blancas.

Para hallar la probabilidad de este suceso, es necesario calcular el número $N(A)$ de resultados, en los cuales aparece el suceso A, $\left(\mathbf{P}(A) = \frac{N(A)}{N}\right)$. Esto se puede hacer del siguiente modo. Separemos mentalmente los dos conjuntos de bolas: r blancas y $n - r$ negras. Las bolas blancas sacadas serán iguales a k, cuando del primer conjunto se elijan k bolas y del segundo, $m - k$ bolas. Designemos a las bolas blancas por $a_1, \ldots, a_r$ y las negras, por $b_1, \ldots, b_{n-r}$; cada elección de bolas blancas $a = (a_{i_1}, \ldots, a_{i_k})$ y de negras $b = (b_{j_1}, \ldots, b_{j_{m-k}})$ puede considerarse como un par (a, b) y el número $N(A)$ coincide con el número de pares diferentes que se pueden formar con distintos elementos a y b. El número total de elementos a, o las distintas elecciones en grupos de k, con las r bolas blancas existentes, es igual al número de combinaciones $C_r^k = \frac{r!}{k!\,(r-k)!}$ y, análogamente, el número total de elementos b es C_{n-r}^{m-k}; por consiguiente, el

número $N(A)$ de pares diferentes (a, b) es igual al producto $C_r^k \cdot C_{n-r}^{m-k}$. Como resultado obtenemos para la probabilidad buscada $\mathbf{P}(A)$ la siguiente expresión:

$$\mathbf{P}(A) = \frac{C_r^k C_{n-r}^{m-k}}{C_n^m}.$$

De hecho, hemos encontrado la distribución de las probabilidades de la magnitud aleatoria ξ, que es igual al número de bolas blancas elegidas al azar de la partida de volumen m (al número de partículas en la zona designada separada del espacio fásico):

$$P_\xi(k) = \frac{C_r^k C_{n-r}^{m-k}}{C_n^m}, \quad k = 0, \ldots, m \tag{1.8}$$

(los parámetros n, r, y m están fijados). Esta es la llamada *distribución hipergeométrica de las probabilidades*.

Para subrayar la importancia de esta distribución examinemos solamente un ejemplo.

Ejemplo (*control por elección de la producción*). Imaginémonos que se tiene una gran partida de piezas, que puede ser admitida o rechazada. Está claro, que no siempre es posible la comprobación de cada una de las piezas (por ejemplo, puede ocurrir que la comprobación haga a la pieza inservible para utilización futura). En estos casos se puede emplear el método de control por elección de la producción. Digamos, el controlador elige, casualmente, de toda la partida m piezas, las comprueba y si el número de las rechazadas entre ellas supera a un número crítico m^* entonces se rechaza toda la partida. ¿Qué ocurre durante tal control, qué partidas y con qué probabilidad son rechazadas? Si en toda la partida hay n piezas (siendo rechazadas en ella justamente r), entonces la probabilidad de que entre las m piezas elegidas al azar resulten ser rechazadas justamente k, se da evidentemente, por la fórmula (1.8) en la que ξ significa el número aleatorio de piezas rechazadas en la partida elegida para el control.

En la tarea inicial sobre la distribución de las partículas en el espacio fásico, se consideran parámetros al número r de partículas existentes y el volumen v de la zona considerada del espacio fásico (del volumen total V). Es incluso más cómodo considerar dados no los números r y V, sino el número medio λ de partículas en la unidad de volumen ($\lambda = r/V$), lo que da la posibilidad, como veremos más adelante, de hallar la distribución de las partículas también en el espacio fásico infinito (más exacto, para $V \to \infty$, $r \to \infty$). Desde este punto de vista la fórmula (1.8) sufre el inconveniente de que en ella figuran «los parámetros sobrantes» n y m. Más adelante se mostrará de hecho (véase también el siguiente § 2) que en las condiciones, cuando la zona considerada del espacio fásico es suficientemente

grande, en relación con la medida de cada una de las partículas, el número ξ de partículas en tal zona tiene, aproximadamente, *la distribución de probabilidades de Poisson*:

$$P_\xi(k) \approx \frac{(\lambda v)^k}{k!} e^{-\lambda v},$$

más exacto, para n, m, $r \to \infty$ y con la constante

$$a = \lambda v = r\frac{v}{V} = r\frac{m}{n}$$

(a es el número medio de partículas en la zona de volumen v), se tiene la siguiente expresión límite para las probabilidades determinadas por la fórmulas (1.8) $P_\xi(k)$ (véase más adelante la fórmula (1.13)):

$$\lim P_\xi(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \ldots \tag{1.9}$$

Volvamos de nuevo a la caja (o como dicen corrientemente a «la urna») con r bolas blancas y $n - r$ negras. En «el esquema de la urna» utilizado por nosotros se eligen m bolas sin retorno. Examinemos un esquema análogo, pero con aquel cambio, que durante la extracción sucesiva de m bolas, la bola sacada cada vez se retorna de nuevo a la urna (e l e c c i ó n c o n r e t o r n o). ¿Cuál será la probabilidad de que entre las m bolas sacadas haya justamente k blancas?

Convengamos en que cada bola de la urna tenga su número. Entonces una elección del volumen m puede ser descrita por la sucesión $(i_1, \ldots, i_m)$, donde i_p significa el número de la bola sacada en el paso p, y durante la elección con retorno para i_p se tienen n valores posibles ($i_p = 1, 2, \ldots, n$). En total habrá n^m elecciones diferentes $(i_1, \ldots, i_m)$ y todas ellas serán equiprobables.

Examinemos las elecciones en las que hay justamente k bolas blancas. Calculemos cuántas elecciones hay $(i_1, \ldots, i_m)$ con bolas blancas en los lugares fijados $m_1, \ldots, m_k$ (es decir, aquellas elecciones en las cuales las bolas $i_{m_1}, \ldots, i_{m_k}$ son blancas y las restantes bolas $m - k$, negras). Para cada una de las bolas blancas $i_{m_1}, \ldots, i_{m_k}$ se tiene r valores posibles (en la urna hay r bolas blancas), y para cada una de las bolas negras hay $n - r$ valores posibles. Por consiguiente, habrá $r^k \cdot (n-r)^{m-k}$ elecciones distintas $(i_1, \ldots, i_m)$ con k bolas blancas en los lugares fijados $m_1, \ldots, m_k$. Pero entre los valores posibles $1, \ldots, m$ podrán existir valores distintos $m_1, \ldots, m_k$ elegidos de C_m^k maneras, de modo que el número total de elecciones $(i_1, \ldots, i_m)$ con m bolas blancas será igual a $C_m^k r^k (n-r)^{m-k}$. Por consiguiente, la probabilidad de que entre las m bolas extraídas haya k blancas será

$$C_m^k \frac{r^k (n-r)^{m-k}}{n^m} = C_m^k p^k (1-p)^{m-k},$$

donde $p = r/n$ coincide con la probabilidad de extraer una bola blanca en cada paso por separado.

Así pues, la magnitud aleatoria ξ que es igual al número de bolas blancas en la elección (con retorno) del volumen m, tiene la distribución de probabilidades

$$P_\xi(k) = C_m^k p^k (1-p)^{m-k}, \quad k = 0, 1, \ldots, m, \qquad (1.10)$$

donde p es la probabilidad de sacar una bola blanca en cada paso por separado. Esta es la tal llamada *distribución binomial*[1]) o como se dice frecuentemente, *la distribución de Bernoulli*.

Está claro, que para un gran número de bolas n y con un volumen pequeño de elección m, la elección sin retorno y la elección con retorno deben dar, prácticamente, un mismo resultado: la probabilidad de que hayan sido extraídas k bolas blancas será aproximadamente, la misma en ambos casos. Esto se reafirma también con la elación límite

$$\frac{C_{n_1}^k \cdot C_{n-n_1}^{m-k}}{C_n^m} = \frac{m!}{k!(m-k)!}\left[\frac{n_1(n_1-1)\ldots(n_1-k+1)}{n(n-1)\ldots(n-k+1)}\right]\times$$

$$\times\left[\frac{n_2(n_2-1)\ldots(n_2-m+k+1)}{(n-k)(n-k-1)\ldots(n-m+1)}\right]\sim\frac{m!}{k!(m-k)!}p^k(1-p)^{m-k},$$

donde $n_1 = r$, $n_2 = n - r$ (compárense las fórmulas (1.8), (1.10)), cuando la parte $p = n_1/n$ de bolas blancas queda constante en la urna, al aumentar ilimitadamente, el número n de bolas (recordemos, que p es la probabilidad de que la bola extraída de la urna al azar sea blanca).

El valor medio $\mathbf{M}\xi$, durante la distribución binomial del número ξ de bolas blancas en la elección de volumen m, será

$$\mathbf{M}\xi = \sum_{k=1}^{m} k\frac{m!}{k!(m-k)!}p^k(1-p)^{m-k} = mp\sum_{j=0}^{m-1}\times$$

$$\times\frac{(m-1)!}{j!(m-1-j)!}p^j(1-p)^{m-1-j} = mp\,[p+(1-p)]^{m-1} = mp. \qquad (1.11)$$

¿Cuál será la distribución de las bolas blancas en la elección de volumen m, cuando la parte $p = r/n$ de bolas blancas disminuye en

---

[1]) $C_m^k = \frac{m!}{k!(m-k)!}$, $k=0, \ldots, m$ son los coeficientes binomiales en el desarrollo de $(p+q)^m = \sum_{k=0}^{m} C_m^k p^k q^{m-k}$ $(q=1-p)$.

la urna ($p \to 0$), pero, en cambio se aumenta en volumen de la elección m ($m \to \infty$) de tal modo, que el valor medio de bolas $a = mp$ quede constante? Tenemos

$$P_\xi(0) = (1-p)^m = \left(1 - \frac{a}{m}\right)^m \to e^{-a},$$

para cada valor fijado $k = 1, 2, \ldots$

$$\frac{P_\xi(k)}{P_\xi(k-1)} = \frac{C_m^k p^k (1-p)^{m-k}}{C_m^{k-1} p^{k-1} (1-p)^{m-k+1}} = \frac{a-(k-1)p}{k(1-p)} \to \frac{a}{k},$$

y por eso

$$\begin{aligned} P_\xi(1) &\to \frac{a}{1!} e^{-a}, \\ P_\xi(2) &\to \frac{a^2}{2!} e^{-a}, \\ &\cdots\cdots\cdots \\ P_\xi(k) &\to \frac{a^k}{k!} e^{-a}, \\ &\cdots\cdots\cdots \end{aligned} \tag{1.12}$$

De tal modo, la magnitud aleatoria ξ tiene en el límite la distribución de probabilidades

$$P_\xi(k) = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \ldots \tag{1.13}$$

Esta es la tal llamada *distribución de Poisson.*

A pesar de que en el límite se trata de hecho de una elección infinita ($m \to \infty$) y la cantidad ξ de «bolas blancas extraídas» puede ser tan grande como se quiera, su valor medio es igual a a:

$$\mathbf{M}\xi = \sum_{k=1}^{\infty} k \frac{a^k}{k!} e^{-a} = a \sum_{k=0}^{\infty} \frac{a^k}{k!} e^{-a} = a. \tag{1.14}$$

Señalemos también que

$$\mathbf{D}\xi = \mathbf{M}(\xi - a)^2 = \mathbf{M}\xi^2 - a^2 = \sum_{k=1}^{\infty} k^2 \frac{a^k}{k!} e^{-a} - a^2 =$$

$$= a \sum_{k=1}^{\infty} \left[(k-1) \frac{a^{k-1} e^{-a}}{(k-1)!} + \frac{a^{k-1}}{(k-1)!} e^{-a} \right] - a^2 = a^2 + a - a^2 = a. \tag{1.15}$$

§ 2. DISTRIBUCION DE POISSON. CORRIENTE UNIFORME DE SUCESOS Y TIEMPO DE ESPERA

1. Distribución de Poisson en las partículas. Volvamos de nuevo al problema de la distribución de las partículas. Imaginémonos que en una zona determinada del espacio V de Euclides se lanzan al azar r partículas, independientemente unas de otras, más exacto, si designamos por ξ_k la situación de la partícula k en el espacio ($k = 1, \ldots, r$), entonces las magnitudes vectoriales aleatorias $\xi_1, \ldots, \xi_r$ son independientes y cada una de ellas tiene la distribución uniforme de probabilidades con la densidad

$$p(x) = \begin{cases} \frac{1}{|V|} & \text{para } x \in V \\ 0 & \text{para las restantes } x, \end{cases} \tag{2.1}$$

donde $|V|$ significa el volumen de la zona V. Designemos por λ al número medio de partículas en la unidad de volumen

$$\lambda = \frac{r}{|V|}.$$

Designemos por $\xi(v)$ el número de partículas, que caen en la zona correspondiente $v \subseteq V$.

Mostremos que en un «espacio fásico» V que se ensancha, ilimitadamente, y a la vez aumenta el número r de partículas (de modo, que el número medio de partículas por unidad de volumen queda, constantemente, igual a λ), la distribución límite de la magnitud aleatoria $\xi(v)$ será la de Poisson

$$\mathbf{P}\{\xi(v) = k\} = \frac{(\lambda |v|)^k}{k!} e^{-\lambda |v|}, \quad k = 0, 1, \ldots, \tag{2.2}$$

donde $|v|$ significa el volumen de la zona v y, aún más, para las zonas que no se cortan $v_1, \ldots, v_n$

$$\mathbf{P}\{\xi(v_1) = k_1, \ldots, \xi(v_n) = k_n\} = \\ = \frac{(\lambda |v_1|)^{k_1} \ldots (\lambda |v_n|)^{k_n}}{k_1! \ldots k_n!} e^{-\lambda(|v_1| + \ldots + |v_n|)}. \tag{2.3}$$

La última relación significa junto con la fórmula (2.2), que para las zonas que no se cortan $v_1, \ldots, v_n$, las magnitudes aleatorias $\xi(v_1), \ldots, \xi(v_n)$ son *independientes*. Esto es intuitivamente claro: durante la existencia de un número infinito de partículas independien-

tes, la caída de tal o cual número finito de partículas en la zona v_1, no influye en el número de partículas que caen en las zonas $v_2, \ldots, v_n$ que no se cortan con v_1.

Pasemos a la deducción de la fórmula (2.3). Cuando están fijados V y r, cada una de las partículas existentes r cae en una de las zonas $v_i \subseteq V$ con las correspondientes probabilidades $p_i = \frac{|v_i|}{|V|}$, $i = 0, 1, \ldots, n$, donde por v_0 habíamos designado al complemento $v_1 \cup v_2 \cup \ldots \cup v_n$ hasta toda la zona V. Dividamos nuestras partículas $\xi_1, \ldots, \xi_r$ en $n+1$ grupos de $k_1, \ldots, k_n, r_0$ partículas en el grupo correspondiente $k_1 + \ldots + k_n = k$, $r_0 = r - k$; el número de variantes distintas en tal división es igual a $\frac{r!}{k_1! \ldots k_n! r_0}$ (véase (1.6)). El suceso $\{\xi(v_1) = k_1, \ldots, \xi(v_n) = k_n\}$ aparece entonces y sólo entonces, cuando un grupo determinado de k_1 partículas cae en la zona v_1, un grupo determinado de k_2 partículas cae en la zona v_2, etc. (las restantes $r_0 = r - k$ partículas caen en la zona v_0). Una partícula tomada por separado, independientemente de las otras partículas, cae en la zona correspondiente v_i con la probabilidad indicada anteriormente p_i, $i = 0, 1, \ldots, n$, y por consiguiente, la probabilidad de que un grupo determinado de k_1 partículas caiga en v_1, otro grupo de k_2 partículas caiga en v_2, etc. será igual al producto $p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_n^{k_n} p_0^{r_0}$. Las distintas variantes de división en los grupos indicados dan $\frac{r!}{k_1! \ldots k_n! r_0!}$ resultados incompatibles, cada uno de los cuales tiene la probabilidad indicada anteriormente $p_1^{k_1} \ldots p_n^{k_n} p_0^{r_0}$ y por consiguiente, el suceso $\{\xi(v_1) = k_1, \ldots, \xi(v_n) = k_n\}$, que coincide con la unión de esos resultados, tiene la probabilidad

$$\frac{r!}{k_1! \ldots k_n! r_0!} p_1^{k_1} \ldots p_n^{k_n} p_0^{r_0}. \tag{2.4}$$

Utilicemos la fórmula de Sterling según la cual

$$r! \sim \sqrt{2\pi r}\, r^r e^{-r}, \quad (r-k)! \sim \sqrt{2\pi (r-k)}\, (r-k)^{r-k} e^{-(r-k)}.$$

Para los valores fijados $k_1, \ldots, k_n$ y $r \to \infty$ tenemos

$$\sqrt{\frac{r}{r-k}} \to 1, \quad \left(\frac{r}{r-k}\right)^{r-k} = \left(1 + \frac{k}{r-k}\right)^{r-k} \to e^k,$$

$$p_i^{k_i} r^{k_i} = \left(\frac{|v_i|}{|V|} r\right)^{k_i} = (\lambda |v_i|)^{k_i}$$

$$i = 1, \ldots, n,$$

$$|v_0| = |V| - (|v_1| + \ldots + |v_n|)$$

y

$$p_0^{r_0} = \left(1 - \frac{|v_1| + \ldots + |v_n|}{|V|}\right)^{r-k} =$$

$$= \left(1 - \frac{\lambda (|v_1| + \ldots + |v_n|)}{r}\right)^{r-k} \to e^{-\lambda(|v_1| + \ldots + |v_n|)}.$$

En total para $r \to \infty$ y el valor constante $\lambda = r/|V|$ tenemos

$$\frac{r!}{k_1! \dots k_n! k_0!} p_1^{k_1} \dots p_n^{k_n} p_0^{k_0} \to \frac{(\lambda |v_1|)^{k_1} \dots (\lambda |v_n|)^{k_n}}{k_1! \dots k_n!} \times \\ \times e^{-\lambda(|v_1|+\dots+|v_n|)},$$

es decir, en el límite obtenemos la fórmula (2.3).

La distribución de Poisson (2.3) se determina por un solo parámetro λ, $0 < \lambda < \infty$, que es el número medio de partículas en la unidad de volumen: $\lambda = \mathbf{M}\xi(v)/|v|$ (véase (1.14)). Esta distribución posee las siguientes propiedades específicas:

a) para las zonas que no se cortan $v_1, \dots, v_n$, las magnitudes aleatorias $\xi(v_1), \dots, \xi(v_n)$ son mutuamente independientes;

b) la probabilidad de caer en la zona v tal o cual número de partículas no depende de la disposición de la zona v en el espacio fásico (sólo depende de su volumen $|v|$);

c) la probabilidad de caer una partícula en la zona de volumen pequeño $|v|$, con exactitud hasta de un infinitamente pequeño de orden superior, es proporcional a $|v|$:

$$\mathbf{P}\{\xi(v) = 1\} = \lambda |v| e^{-\lambda|v|} = \lambda |v| + o(|v|),$$

y la probabilidad de caer más de una partícula en la zona de volumen $|v|$ es una cantidad infinitamente pequeña de orden superior para $|v| \to 0$:

$$\mathbf{P}\{\xi(0) > 1\} = \sum_{k=2}^{\infty} \frac{(\lambda |v|)^k}{k!} e^{-\lambda|v|} = o(|v|).$$

Mostremos, que *para cualquier distribución de las partículas, que satisfaga a las condiciones* a), b) y c) *indicadas, anteriormente, es obligatoria la distribución de Poisson.*

En esencia, sólo hay que mostrar que tiene lugar la fórmula (2.2) (de donde para la condición a) se deduce también la fórmula más general (2.3)). Utilicemos el mismo modo de división de la zona v en pequeñas celdas Δ_k, $k = 1, \dots, n$ (de volumen $|\Delta_k| = \frac{|v|}{n}$), como en el punto 2 del párrafo anterior (compárese con (1.9) y demás). Debido a las condiciones a) y b), la probabilidad del suceso

$$\{\xi(\Delta_{i_p}) = 1, \quad p = 1, \dots, k, \quad \xi(\Delta_{i_p}) = 0, \\ p = k+1, \dots, n\}$$

es una misma para todas las combinaciones $(i_1, \dots, i_k)$, $(i_{k+1}, \dots, i_n)$ y es igual al producto

$$(\mathbf{P}\{\xi(\Delta_1) = 1\})^k (\mathbf{P}\{\xi(\Delta_1) = 0\})^{n-k};$$

y debido a la condición c), para $n \to \infty$ la probabilidad de que caiga más de una partícula en cualquier celda no supera a la suma

$$\sum_{k=1}^{n} \mathbf{P}\{\xi(\Delta_k) > 1\} = n \cdot o\{|\Delta_k|\} \to 0,$$

y por consiguiente, suponiendo que

$$p = \mathbf{P}\{\xi(\Delta_1) = 1\} = \lambda \frac{|v|}{n} + o\left(\frac{1}{n}\right),$$

para $n \to \infty$ obtenemos (véase (1.13))

$$\mathbf{P}\{\xi(v) = k\} \sim \qquad 1104\ 168\ 2$$

$$\sim \sum_{(i_1, \ldots, i_k)} \mathbf{P}\{\xi(\Delta_{i_p}) = 1,\ p = 1, \ldots k,\ \xi(\Delta_{i_p}) = 0,\ p = k+1, \ldots, n\} =$$

$$= C_n^k (\mathbf{P}\{\xi(\Delta_1) = 1\})^k (\mathbf{P}\{\xi(\Delta_1) = 0\})^{n-k} \sim$$

$$\sim C_n^k p^k (1-p)^{n-k} \sim \frac{a^k}{k!} e^{-a},$$

donde $a = np = \lambda |v|$.

Señalemos además la siguiente característica de la distribución de Poisson: *bajo la condición de que caigan justamente r partículas en una determinada zona V, su distribución dentro de esta zona será tal, como si las hubiesen tirado al azar en V, independientemente una de otra; más exactamente, las magnitudes vectoriales* $\xi_1, \ldots, \xi_r$, *que describen la posición de estas partículas, serán independientes y estarán distribuidas uniformemente en la zona V (con una densidad de probabilidad de la forma* (2.1)).

En efecto, para cualesquiera zonas que no se cortan $v_1, \ldots, v_n$ la probabilidad convencional $\mathbf{P}\{\xi(v_1) = k_1, \ldots, \xi(v_n) = k_n \mid \xi(V) = r\}$ es

$$\frac{\mathbf{P}\{\xi(v_1) = k_1, \ldots, \xi(v_n) = k_n,\ \xi(v_0) = r_0\}}{\mathbf{P}\{\xi(V) = r\}} =$$

$$= \frac{\frac{(\lambda|v_1|)^{k_1}}{k_1!} \cdots \frac{(\lambda|v_n|)^{k_n}}{k_n!} e^{-\lambda(|v_1|+\ldots+|v_n|+|v_0|)}}{\frac{(\lambda|V|)^r}{r!} e^{-\lambda|V|}} =$$

$$= \frac{r!}{k_1! \ldots k_n! r_0!} p_1^{k_1} \cdots p_n^{k_n} p_0^{r_0},$$

lo que demuestra nuestra afirmación (compárese con (2.4)).

Ejemplo (proceso de la desintegración radiactiva). Como es conocido, el radio (Ra) con el transcurso del tiempo se transforma en radón (Rn). El núcleo del átomo de Ra que se desintegra emite la llamada partícula α (núcleo del átomo de helio He). La desintegración de un átomo por separado de Ra se realiza independientemente del estado de otros átomos de la substancia y la emisión de rayos α se representa

en sí a un flujo de una gran cantidad de partículas independientes. Es natural esperar, que la distribución de las partículas α en el tiempo será la distribución de Poisson; la probabilidad de que el número $\xi(\Delta)$ de partículas α emitidas sea igual a k durante el intervalo de tiempo Δ, deberá ser

$$\mathbf{P}\{\xi(\Delta)=k\}=\frac{\lambda|\Delta|^k}{|k!|}e^{-\lambda|\Delta|},\quad k=0,\ 1,\ \ldots, \tag{2.5}$$

donde $|\Delta|$ es la amplitud del intervalo de tiempo examinado, y λ es el número medio de partículas α emitidas en la unidad de tiempo: $\lambda=\frac{M\xi(\Delta)}{\Delta}$ (el proceso de desintegración radioactiva será examinado más detalladamente en el siguiente punto 2). Este resultado se confirma bien con los datos experimentales[1]).

Naturalmente, la distribución de Poisson se origina no sólo al examinar el flujo uniforme de partículas independientes. Digamos, se puede tratar de una corriente homogénea de exigencias que se reciben en un determinado sistema de servicio (por ejemplo, al surtidor de gasolina llegan los autos, se reciben peticiones en el buró de información, en la estación telefónica se registra un flujo de fallos en los abonados, etc.). En general, la cuestión puede tratar sobre una corriente homogénea de sucesos independientes, registrados en el tiempo: los momentos de aparición de estos sucesos se puede interpretar como «partículas» distribuidas aleatoriamente sobre la recta real (eje del tiempo). Si en el intervalo de tiempo Δ se cumplen, para el número de sucesos $\xi(\Delta)$, las condiciones descritas anteriormente a), b) y c):

a) para los intervalos de tiempo que no se cortan $\Delta_1, \ldots, \Delta_n$, las magnitudes aleatorias $\xi(\Delta_1), \ldots, \xi(\Delta_n)$ serán independientes;

b) la probabilidad de tal o cual suceso en el intervalo Δ no depende del origen de recuento del tiempo (no depende de la posición de Δ en el eje de tiempo);

c) la probabilidad de aparición del suceso en un pequeño intervalo de tiempo Δ es proporcional a la amplitud $|\Delta|$ de este intervalo, y la probabilidad de aparición de más de un suceso, tiene un orden infinitesimal superior en comparación con $|\Delta|$, entonces a tal corriente de sucesos se la llama *de Poisson* y la distribución de probabilidades $\xi(\Delta)$ se describe por la fórmula (2.5).

2. Tiempo de espera del suceso aleatorio. Examinemos los lanzamientos sucesivos de la moneda hasta la primera caída «cruz», considerando que la moneda se lanza una vez en la unidad de tiempo. Está claro que la duración de la prueba (tiempo de espera de la «cruz») es aleatorio casual, con ello, debido a la independencia de las pruebas por separado (lanzamiento de la moneda) en cualquier momento

[1]) Véase, por ejemplo, el libro citado anteriormente de W. Feller.

$t < 0$, la posición del observador no será nada mejor (en sentido del tiempo de espera de «cruz») que en el momento inicial $t = 0$: la probabilidad de esperar después del momento t un tiempo más s, es la misma que esperar este mismo tiempo s a partir del momento inicial; más exactamente, si τ es el tiempo de espera, entonces para la condición $\tau > t$ deberá ser

$$\mathbf{P}(\tau > t + s \mid \tau > t) = \mathbf{P}(\tau > s). \tag{2.6}$$

En efecto, para cualquier $t = 0, 1, \ldots$, el suceso $\{\tau > t\}$ significa que en t pruebas independientes no caerá ni una vez «cruz» y

$$\mathbf{P}\{\tau > t\} = (1 - p)^t, \quad t = 0, 1, \ldots, \tag{2.7}$$

donde $p = 1/2$ es la probabilidad de que caiga «cruz» en cada prueba por separado, por eso

$$\mathbf{P}\{\tau > t + s \mid \tau > t\} = \frac{\mathbf{P}\{\tau > t + s\}}{\mathbf{P}\{\tau > t\}} = (1 - p)^s = \mathbf{P}\{\tau > s\}.$$

La relación (2.6), considerada por sí misma, es característica para el tiempo de espera τ de tal suceso A, cuya espera hasta el momento t de ningún modo aproxima, prácticamente, su aparición. En esta situación, en el ejemplo más elemental que fue expuesto anteriormente, la distribución de la magnitud aleatoria τ será tal que

$$\mathbf{P}\{\tau > t\} = e^{-\lambda t}, \quad t \geqslant 0, \tag{2.8}$$

donde λ es un determinado parámetro ($\lambda \geqslant 0$); para el valor discreto de $t = 0, 1, \ldots$, la fórmula (2.8) puede ser escrita en la forma (2.7) si ponemos $p = 1 - e^{-\lambda}$.

Para el valor discreto de t esta distribución de probabilidades se llama distribución *geométrica*:

$$P_\tau(k) = e^{-\lambda(k-1)} - e^{-\lambda k} = p(1 - p)^{k-1}, \quad k = 1, 2, \ldots \tag{2.9}$$

Para el valor continuo de t, $0 \leqslant t < \infty$, la distribución (2.8) tiene una densidad de probabilidad

$$p_\tau(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geqslant 0, \\ 0 & \text{para } t < 0 \end{cases} \tag{2.10}$$

(ya que la función de distribución τ es $F_\tau(t) = 1 - \mathbf{P}\{\tau > t\}$ y $p_\tau(t) = F'_\tau(t)$); tal distribución se llama *exponencial*. El parámetro indicado λ tiene un sentido probabilístico sencillo, es decir, $1/\lambda$ es el tiempo medio de espera

$$\frac{1}{\lambda} = \mathbf{M}\tau = \int_0^\infty t p_\tau(t)\, dt. \tag{2.11}$$

Deduzcamos la fórmula (2.8) de la relación (2.6) suponiendo que $\tau = 0$, sólo cuando la probabilidad es igual a cero, y que en caso del tiempo continuo, la magnitud τ tiene para $t \geqslant 0$ una densidad de probabilidad continua. Determinemos la función $f(t)$ por la igualdad $f(t) = \mathbf{P}\{\tau > t\}$; según la fórmula de la probabilidad total

$$\mathbf{P}\{\tau > t + s\} = \mathbf{P}\{\tau > s\} \cdot \mathbf{P}\{\tau > t\},$$

y, por consiguiente, la función $f(t)$ será tal, que para cualesquiera $s,\ t \geqslant 0$

$$f(s + t) = f(s) f(t),$$

o bien

$$\log f(s + t) = \log f(s) + \log f(t),$$

siendo $f(0) = \mathbf{P}\{\tau > 0\} = 1$. Para el valor discreto de $t = 0, 1, \ldots$ obtenemos de aquí que $\log f(t)$, $\log f(0) = 0$ es una función lineal:

$$\log f(t) = t \log f(1)$$

y

$$f(t) = (1 - p)^t = e^{-\lambda t}, \quad \text{donde} \quad p = 1 - f(1) = \mathbf{P}\{\tau = 1\}.$$

Para el tiempo continuo t (en condiciones de que existe la densidad $p_\tau(t) = -f'(t + 0)$, $t \geqslant 0$) la diferenciación por s nos da

$$\frac{f'(t+s)}{f(t+s)} = \frac{f'(s)}{f(s)},$$

de donde para $s = 0$ obtenemos

$$\frac{f'(t)}{f(t)} = -\lambda,$$

donde $-\lambda = f'(0) \leqslant 0$, ya que $f(0) = 1$ es el valor máximo de la función $f(t) = \mathbf{P}\{\tau > t\}$, $t \geqslant 0$. Por consiguiente, $f(t) = e^{-\lambda t}$, lo que se exigía demostrar.

Ejemplo (modelo de la desintegración radiactiva). Anteriormente ya recordamos que la distribución de las partículas α emitidas por el radio es la distribución de Poisson. Supongamos, que la transformación del radio Ra en radón Rn (en la que se emiten partículas α) se realiza de tal modo que en el intervalo de tiempo (t_0, t_1) un átomo de Ra tomado por separado se transforma en un átomo de Rn con una determinada probabilidad $p = p(t)$, dependiente de la amplitud $t = t_1 - t_0$ del intervalo examinado.

Si hasta el momento t_1 no ocurrió la transición Ra → Rn, entonces en el nuevo momento inicial t_1 trataremos del mismo átomo de Ra y la transición Ra → Rn en el tiempo sucesivo $s = t_2 - t_1$, de acuerdo a nuestra suposición, deberá realizarse con la correspondiente probabilidad $p(s)$. Evidentemente, el tiempo τ de espera de la desintegración Ra → Rn del átomo dado de Ra (empezando desde el momento

t_0) satisface a la relación (2.6), ya que la probabilidad condicional $\mathbf{P}\{\tau > t+s \mid \tau > t\}$ es la probabilidad de que el átomo Ra conservándose hasta el momento t_1, no se desintegra tampoco en el siguiente intervalo de tiempo (t_1, t_2), y por suposición, esta probabilidad (así como la probabilidad $\mathbf{P}\{\tau > s\}$) es igual a $1 - p(s)$, $s = t_2 - t_1$. Por consiguiente, el tiempo de espera τ tiene la distribución exponencial de probabilidades (2.8).

Aclaremos el sentido físico de la constante correspondiente λ. Ya sabemos, que $1/\lambda$ es el tiempo medio de espera de la desintegración Ra → Rn para un átomo por separado (véase (2.11)), es natural considerar, *que la constante λ es una misma para todos los átomos de Ra.*

Designemos por n_0 la cantidad de radio (digamos, el número de *átomos de Ra) en el momento inicial* t_0. Cada átomo de Ra por separado se desintegra (Ra → Rn) en el tiempo t sucesivo con la probabilidad

$$p(t) = 1 - e^{-\lambda t},$$

y si designamos por $\xi(t)$ al número de átomos Ra desintegrados en este tiempo (que coincide con el número de partículas α emitidas) entonces el valor medio $\mathbf{M}\xi(t)$ será

$$\mathbf{M}\xi(t) = n_0 p(t) = n_0 (1 - e^{-\lambda t}).$$

Por consiguiente, *la cantidad restante de radio después del tiempo* t será en su media

$$n(t) = \mathbf{M}(n_0 - \xi(t)) = n_0 - \mathbf{M}\xi(t) = n_0 e^{-\lambda t}.$$

La dependencia exponencial de t puede ser caracterizada no sólo por el exponente λ, sino también por la llamada constante T de semidesintegración, definida como el tiempo durante el cual se desintegra exactamente, la mitad de la substancia inicial; para hallar T tenemos *la ecuación*

$$n(T) = \frac{n_0}{2},$$

que da el valor $T = \dfrac{\log 2}{\lambda}$ (*para el radio ha sido hallado experimentalmente* $T = 1590$ años).

Examinemos ahora un modelo general de la corriente de sucesos de Poisson, en el cual se supone que, independientemente de cuándo y *cuántos sucesos se realizaron hasta el momento corriente* t_0, en el intervalo de tiempo (t_0, t_1) aparecen exactamente k sucesos con la correspondiente probabilidad $\dfrac{(\lambda t)^k}{k!} e^{-\lambda t}$, $k = 0, 1, \ldots$ (donde $t = t_1 - t_0$).

Vemos que *el tiempo aleatorio de espera (después del momento* t_0*) del suceso de turno tiene una distribución exponencial de probabilidades*

con el parámetro λ:

$$\mathbf{P}\{\tau_1 - t_0 > t\} = \mathbf{P}\{\xi(t_0, t_1) = 0\} = e^{-\lambda t}, \qquad (2.12)$$

donde τ_1 significa el momento de aparición del suceso (el primero después del momento t_0), y $\xi(t_0, t_1)$ es el número de sucesos en el intervalo de tiempo (t_0, t_1); el parámetro λ de esta distribución exponencial es el mismo que en la corriente inicial de Poisson (λ es el número medio de sucesos ocurridos en la unidad de tiempo).

Por consiguiente, en la corriente de sucesos de Poisson, incluso después de una larga espera del suceso de turno, éste se realiza en el siguiente intervalo de tiempo (t_1, t_2) con la misma probabilidad $1 - e^{-\lambda s}$, $s = t_2 - t_1$, como si no hubiese habido nunca ninguna espera y t_1, fuese el momento inicial de observación ($t_1 = t_0$):

$$\mathbf{P}\{\tau_1 - t_0 > t + s \mid \tau_1 - t_0 > t\} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbf{P}\{\tau_1 - t_0 > s\}.$$

Supongamos que $\tau_1, \ldots, \tau_n$ son los momentos aleatorios, en que se realizan, siguiendo uno a otro, los sucesos de la corriente de Poisson examinada. Tomemos un momento arbitrario t_1, $\tau_1 \leqslant t_1 < \tau_2$. Después del momento t_1, el suceso de turno aparece, independiente de τ_1, durante el tiempo t con la probabilidad $1 - e^{-\lambda t}$, y por eso, la probabilidad condicional de que $\{\tau_2 - t_1 \leqslant t\}$ para cualquier τ_1, $\tau_1 \leqslant t_1$ es

$$\mathbf{P}\{\tau_2 - t_1 \leqslant t \mid \tau_1,\ \tau_1 \leqslant t_1\} = 1 - e^{-\lambda t}.$$

De aquí, para $\tau_1 = t_1$ obtenemos

$$\mathbf{P}\{\tau_2 - t_1 \leqslant t \mid \tau_1 = t_1\} = \mathbf{P}\{\tau_2 - \tau_1 \leqslant t \mid \tau_1 = t_1\} = 1 - e^{-\lambda t}.$$

Se ve que la densidad condicional de probabilidad $p(t \mid t_1)$, $t \geqslant 0$ de la magnitud $\tau_2 - \tau_1$ a condición de que $\tau_1 = t_1$ para todos los $t_1 \geqslant t_0$ es una misma e igual a $\lambda e^{-\lambda t}$. Por consiguiente, (según la fórmula (3.1)) la densidad «incondicional» es

$$p(t) = \int_{t_0}^{\infty} p(t \mid t_1)\, p_{\tau_1}(t_1)\, dt_1 = \lambda e^{-\lambda t},$$

es decir, *la amplitud del intervalo de tiempo* $\tau_2 - \tau_1$ *tiene también la distribución exponencial con el mismo parámetro* λ:

$$\mathbf{P}\{\tau_2 - \tau_1 > t\} = e^{-\lambda t}, \quad t \geqslant 0. \qquad (2.13)$$

Consideraciones análogas, referentes a los momentos τ_{n-1} y τ_n, permiten concluir que independientemente de los momentos de aparición de los sucesos anteriores $\tau_1, \tau_2, \ldots, \tau_{n-1}$ (lo que significa

también que, independientemente de los intervalos $\tau_2 - \tau_1, \ldots$ $\ldots, \tau_{n-1} - \tau_{n-2}, \tau_1 - t_0$)

$$\mathbf{P}\{\tau_n - \tau_{n-1} > t \mid \tau_1, \tau_2, \ldots, \tau_{n-1}\} = \mathbf{P}\{\tau_n - \tau_{n-1} > t\} = e^{-\lambda t}. \quad (2.14)$$

Como resultado llegamos a la siguiente conclusión: *los intervalos* $\tau_1 - t_0$, $\tau_2 - \tau_1, \ldots$, *que separan los sucesos de la corriente de Poisson que aparecen sucesivamente, son magnitudes aleatorias independientes, que tienen la misma distribución exponencial con el parámetro* λ (λ *es el número medio de sucesos realizados en la unidad de tiempo*).

Sean $\xi_1, \xi_2, \ldots$ magnitudes aleatorias independientes que tienen la misma distribución exponencial con parámetro λ; supongamos que

$$S_n = \xi_1 + \ldots + \xi_n. \quad (2.15)$$

Si $\xi_1 = \tau_1 - t_0$, $\xi_2 = \tau_2 - \tau_1, \ldots$ son los intervalos en los sucesos de la corriente de Poisson, entonces la magnitud aleatoria S_n es igual al tiempo de espera del suceso n según el cálculo: $S_n = \tau_n - t_0$. Evidentemente, el suceso $\{S_n \leqslant t\}$ significa que en el tiempo t por lo menos se realizó n sucesos y, por consiguiente, la función de la distribución de probabilidades de la magnitud no negativa S_n es

$$\mathbf{P}\{S_n \leqslant t\} = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad t \geqslant 0.$$

Diferenciando esta expresión por t, hallamos la densidad de probabilidad correspondiente, que tiene la forma

$$p(t) = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geqslant 0 \quad (2.16)$$

($p(t) = 0$ para $t < 0$).

La distribución de probabilidades obtenida entra en la familia de las tal llamadas «*distribuciones gamma*» (en adelante nos pondremos en conocimiento con algunas de ellas). La densidad general de la «distribución gamma» se da por la fórmula

$$p(t) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} & \text{para } t \geqslant 0, \\ 0 & \text{para } t < 0, \end{cases} \quad (2.17)$$

donde α, β son parámetros positivos, y

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt \quad (2.18)$$

$$(\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma(n+1) = n!)$$

es la llamada función gamma de Euler. La distribución (2.16) se obtiene para $\alpha = n$, $\beta = \lambda$; para $n = 1$ la cuestión se refiere a la distribución exponencial.

Como conclusión de este punto nos detendremos en un hecho, que puede parecer paradójico a primera vista. Precisamente, examinemos, como anteriormente, la corriente de sucesos de Poisson: independientemente, de cuándo y de cuántos sucesos se realizaron (o se realizarán) fuera de un determinado intervalo de tiempo (s, t), en este intervalo ocurren k sucesos con la correspondiente probabilidad

$$\frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)}, \quad k = 0, 1, \ldots$$

Supongamos que los sucesos se registran en el eje de tiempo $-\infty < t < \infty$ y el observador, que llegó en determinado momento t_0, se interesa por los intervalos entre el suceso anterior (el momento de su aparición lo designamos por τ_{-1}) y el suceso siguiente cuyo momento de aparición lo designamos por τ_1, como anteriormente.

Nosotros establecimos que los intervalos (digamos $\tau_2 - \tau_1$, $\tau_3 - \tau_2$, etc.) entre los sucesos consecutivos, lo mismo que la magnitud $\tau_1 - t_0$, tienen una distribución exponencial con el parámetro λ. Se podría pensar que la diferencia $(\tau_1 - \tau_{-1})$ tiene la misma distribución, aunque, surge en seguida la incomprensión debido a la desigualdad $(\tau_1 - \tau_{-1}) > (\tau_1 - t_0)$, de la cual se deduce, que las magnitudes indicadas no pueden tener una misma distribución de probabilidades. Se puede abordar la solución del problema sobre la distribución de la magnitud $(\tau_1 - \tau_{-1})$, teniendo en cuenta la circunstancia de que la dirección del transcurso del tiempo no influye en nada sobre las leyes probabilísticas de la corriente de sucesos de Poisson, ellas serán también las mismas para la corriente de sucesos «invertida hacia atrás» (formalmente, se obtiene con el cambio de la variable t por $-t$, $-\infty < t < \infty$). Por consiguiente, para cualquier momento t_0 fijado, de modo que $\tau_{-1} < t_0 < \tau_1$ no sólo, la magnitud $(\tau_1 - t_0)$, que es el tiempo de espera del suceso de turno, tendrá la distribución exponencial de probabilidades, sino que también deberá poseer esta propiedad la magnitud $(t_0 - \tau_{-1})$, que es el tiempo de espera del suceso de turno en la «corriente invertida». Además de esto, como ya indicamos, la magnitud $(\tau_1 - t_0)$ no depende de cuándo se realizó, precisamente, el suceso anterior al momento t_0 y de tal modo, para la condición $\tau_{-1} < t_0 < \tau_1$ las magnitudes $(\tau_1 - t_0)$ y $(t_0 - \tau_{-1})$ serán independientes. Por consiguiente, la distribución de la diferencia $(\tau_1 - \tau_{-1})$, que es el intervalo entre los sucesos vecinos (a condición de que $\tau_{-1} < t_0 < \tau_1$, para un determinado momento fijado t_0), será la misma, que en la suma $(\tau_1 - t_0) + (t_0 - \tau_{-1})$ de dos magnitudes independientes, que tienen la misma distribución exponencial (la densidad de probabilidad correspondiente se da por la fórmula (2.16) para $n = 2$).

§ 3. PRUEBAS DE BERNOULLI. MOVIMIENTO BROWNIANO Y DISTRIBUCIONES DE PROBABILIDADES LIGADAS A ESTE

1. Pruebas de Bernoulli y distribución binomial. Aproximaciones de Poisson y normal. Las pruebas iguales e independientes entre sí, en cada una de las cuales se examina un determinado suceso A, que aparece con una misma probabilidad positiva $p = \mathbf{P}(A)$, se denominan *pruebas de Bernoulli*. El propio suceso A se llama, convencionalmente, «éxito» y el suceso complementario $\overline{A}$ que aparece en cada una de las pruebas examinadas con la probabilidad $q = 1 - p$, convencionalmente se llama «fallo».

Designemos al «éxito» en la prueba por separado con la unidad (1), y al «fallo» con el cero (0). Entonces cada resultado elemental $\omega = (\omega_1, \ldots, \omega_n)$ de n pruebas da la sucesión correspondiente de ceros y unidades ($\omega_i = 1$ con la probabilidad p y $\omega_i = 0$ con la probabilidad $1 - p$). La probabilidad $\mathbf{P}(\omega)$ del resultado elemental ω, para el cual aparece justamente k veces el «éxito» y $n - k$ veces aparece el «fallo», debido a la independencia de las pruebas por separado, es

$$\mathbf{P}(\omega) = p^k q^{n-k}.$$

Examinemos la magnitud aleatoria ξ, igual al número total de «éxitos» en n pruebas de Bernoulli: $\xi(\omega) = k$, si aparece justamente k veces el «éxito» en el resultado elemental ω.

El número de resultados distintos ω, que conducen a un mismo número k de «éxitos», es igual al número de combinaciones de n tomados en grupos de k, que alcanza a $C_n^k = \frac{n!}{k!\,(n-k)!}$. Todos estos resultados ω tienen una misma probabilidad $\mathbf{P}(\omega) = p^k q^{n-k}$, de modo que el suceso $\{\xi = k\}$ tiene la probabilidad $C_n^k p^k q^{n-k}$. Por lo tanto, la distribución de las probabilidades de la magnitud aleatoria ξ se da por la fórmula

$$P_\xi(k) = C_n^k p^k q^{n-k}, \quad k = 0, \ldots, n. \tag{3.1}$$

Esta es *la distribución binomial* (fig. 12), que ya apareció antes (véase (1.10)) al examinar el esquema de la urna, en el cual el «éxito» era la extracción de la bola blanca, y el «fallo» la extracción de la bola negra. Ella se da por dos parámetros: por la probabilidad de un «éxito» por separado p y por el número n de pruebas.

Es útil indicar, que la magnitud aleatoria ξ es la suma de n magnitudes *independientes* $\xi_1, \ldots, \xi_n$ que se determinan del siguiente

modo: $\xi_k = 1$ si aparece el «éxito» en la prueba k-ésima y $\xi_k = 0$, si en esta prueba aparece el «fallo»:

$$\xi = \xi_1 + \ldots + \xi_n. \tag{3.2}$$

Si observamos que

$$M\xi_k = p,$$
$$D\xi_k = M\xi_k^2 - (M\xi_k)^2 = p - p^2 = p(1-p) = pq,$$

entonces utilizando (3.2), obtenemos para la esperanza matemática y la dispersión de la magnitud aleatoria ξ la siguiente expresión:

$$M\xi = np, \quad D\xi = npq. \tag{3.3}$$

Para un gran número n de pruebas y una probabilidad p relativamente pequeña, cuando cada unó de los éxitos es un suceso relativa-

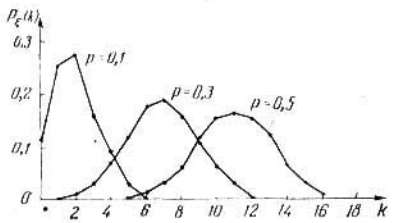

Fig. 12. Los puntos en los gráficos significan las probabilidades binomiale $P_\xi(k) = C_n^k p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$, para $n = 20$, que responden a parámetro $p = 0, 1, 0{,}3, 0{,}5$, respectivamente.

mente raro, pero el número medio de «éxitos» np es suficientemente grande, se puede considerar que

$$P_\xi(k) \approx \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \ldots, \tag{3.4}$$

donde $a = np$ es el número medio de éxitos (véase (1.12). Esta es la tal llamada *aproximación de Poisson para la distribución binomial (fig. 13)*.

En las aplicaciones, las pruebas de Bernoulli sirven, frecuentemente, para la determinación experimental de la probabilidad

$$p = P(A)$$

de tal o cual suceso A (del «éxito» en una prueba por separado) para cuya valuación empírica se toma la frecuencia observada del «éxito»

$$p \approx \frac{n(A)}{n} \left(= \frac{1}{n} \sum_{k=1}^{n} \xi_k \right). \tag{3.5}$$

Debido a la ley de los grandes números

$$p = \lim_{n\to\infty} \frac{n(A)}{n},$$

con ello, la desviación media cuadrática de la valuación $\frac{n(A)}{n}$ de la probabilidad desconocida $\mathbf{P}(A)$, tomada por el observador, desde su

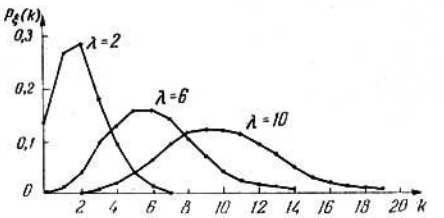

Fig. 13. Los puntos en los gráficos significan las probabilidades de Poisson $P_\xi(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \ldots$, que responden al parámetro $\lambda = 2, 6, 10$, respectivamente ($\lambda = np$ para $n = 20$ y $p = 0{,}1,\ 0{,}3,\ 0{,}5$; compárese con la fig. 12).

verdadero valor p, llega hasta

$$\left\| \frac{n(A)}{n} - p \right\| = \frac{\sqrt{\mathbf{D}S_n}}{n} = \sqrt{\frac{pq}{n}},$$

donde $S_n = \sum_{k=1}^{n} \xi_k$ sólo es la nueva designación para el número de «éxitos» en las n pruebas independientes. De tal modo en la representación

$$\frac{n(A)}{n} - p = \sqrt{\frac{pq}{n}} \cdot S_n^*$$

la magnitud aleatoria $S_n^* = \frac{S_n - \mathbf{M}S_n}{\sqrt{\mathbf{D}S_n}}$ tiene el valor medio 0 y la dispersión 1: se ve que la desviación probable de la frecuencia $\frac{n(A)}{n}$ del valor desconocido de p es del orden $\frac{1}{\sqrt{n}}$. Más exacto es el resultado siguiente.

Teorema (de Moivre—Laplace). *Para $n \to \infty$ la magnitud aleatoria S_n^* tiene una distribución de probabilidades límite, a saber*

$$\lim_{n\to\infty} \mathbf{P}\{x' \leqslant S_n^* \leqslant x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2}\, dx \qquad (3.6)$$

para cualesquiera valores fijados x', x'', $(x' \leqslant x'')$.

Esta es la llamada *distribución de probabilidades normal* (o *de Gauss*) con la densidad

$$p(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\quad -\infty<x<\infty. \tag{3.7}$$

Pasemos a la demostración de la relación límite (3.6). La magnitud normalizada S_n^*, según el caso, toma uno de los valores x de la forma

$$x=\frac{k-np}{\sqrt{npq}},\quad k=0,\ 1,\ \ldots,\ n,$$

que dividen los segmentos $\left[\frac{-np}{\sqrt{npq}},\ \frac{nq}{\sqrt{npq}}\right]$ en intervalos iguales de longitud

$$\Delta x=\frac{1}{\sqrt{npq}},$$

siendo

$$\mathbf{P}\{S_n^*=x\}=P_{S_n}(k)=\frac{n!}{k!\,(n-k)!}p^kq^{n-k},\quad k=0,\ 1,\ \ldots,\ n.$$

Evidentemente, para $n\to\infty$

$$k=np+\sqrt{npq}\,x\to\infty,\quad n-k=nq-\sqrt{npq}\,x\to\infty$$

uniformemente según x, $x'\leqslant x\leqslant x''$. Utilizando la fórmula de Sterling obtenemos, que

$$P_{S_n}(k)\sim\frac{\sqrt{2\pi n}\,n^ne^{-n}}{\sqrt{2\pi k}\,k^ke^{-k}\sqrt{2\pi(n-k)}\,(n-k)^{n-k}e^{-(n-k)}}=$$
$$=\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$$

Luego,

$$\frac{k}{np}=1+\sqrt{\frac{q}{np}}\,x,\quad \frac{n-k}{nq}=1-\sqrt{\frac{p}{nq}}\,x$$

y utilizando el desarrollo

$$\ln(1+\alpha_n)\sim\alpha_n-\tfrac{1}{2}\alpha_n^2$$

(para $\alpha_n\to 0$) obtenemos que

$$\ln\left(\frac{k}{np}\right)^{-k}=-k\ln\left(1+\sqrt{\frac{q}{np}}\,x\right)\sim-(np+\sqrt{npq}\,x)\times$$
$$\times\left[\sqrt{\frac{q}{np}}\,x-\frac{1}{2}\frac{q}{np}x^2\right];$$
$$\ln\left(\frac{n-k}{nq}\right)^{-(n-k)}=-(n-k)\ln\left(1-\sqrt{\frac{p}{nq}}\,x\right)\sim$$
$$\sim-(nq-\sqrt{npq}\,x)\left[-\sqrt{\frac{p}{nq}}\,x-\frac{1}{2}\frac{p}{nq}x^2\right]$$

Sumando estas expresiones llegamos a la siguiente relación:

$$\lim_{n\to\infty} \ln\left[\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}\right] = -\frac{1}{2}x^2,$$

y, por consiguiente,

$$\lim_{n\to\infty} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} = e^{-x^2/2}.$$

Luego,

$$\sqrt{\frac{n}{k(n-k)}} \sim \sqrt{\frac{n}{npnq}} = \frac{1}{\sqrt{npq}} = \Delta x,$$

y como resultado obtenemos, que

$$\mathbf{P}\{S_n^* = x\} \sim \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \Delta x, \quad \Delta x = \frac{1}{\sqrt{npq}}$$

siendo uniforme por todos los $x = \frac{k-np}{\sqrt{npq}}$, $x' \leqslant x \leqslant x''$. Por consiguiente,

$$\mathbf{P}\{S_n^* = x\} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \Delta x + o(\Delta x)$$

uniforme por todos los $x = \frac{k-np}{\sqrt{npq}}$, $x' \leqslant x \leqslant x''$ y para $n \to \infty$

$$\mathbf{P}\{x' \leqslant S_n^* \leqslant x''\} = \sum_{x'}^{x''} \mathbf{P}\{S_n^* = x\} = \frac{1}{\sqrt{2\pi}} \times$$

$$\times \sum_{x'}^{x''} [e^{-x^2/2}\Delta x + o(\Delta x)] = \frac{1}{\sqrt{2\pi}} \times$$

$$\times \sum_{x'}^{x''} e^{-x^2/2}\Delta x + o(1) \sim \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx,$$

lo que se exigía demostrar.

De tal modo, para la distribución de probabilidades de la magnitud aleatoria S_n^* tenemos la siguiente aproximación (normal):

$$\mathbf{P}\{x' \leqslant S_n^* \leqslant x''\} \approx \Phi(x'') - \Phi(x'),$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$ es la función de distribución de probabilidades normal (con la densidad (3.7)) (fig. 14). Una aproximación algo más exacta de la fórmula [1]) es

$$\mathbf{P}\{x' \leqslant S_n^* \leqslant x''\} \approx \Phi\left(x'' + \frac{\Delta x}{2}\right) - \Phi\left(x' - \frac{\Delta x}{2}\right),$$

[1]) Véase, por ejemplo, el libro citado anteriormente de V. Feller.

donde $\Delta x = \frac{1}{\sqrt{npq}}$. Para la diferencia de las funciones de distribuciones $F(x) = \mathbf{P}\{S_n^* \leqslant x\}$ y $\Phi(x)$ se tiene la siguiente valuación (véase más adelante (4.19)):

$$\sup_x |F(x) - \Phi(x)| \leqslant C \frac{1}{\sqrt{n}}, \qquad (3.8)$$

donde $C \leqslant \frac{p^2+q^2}{\sqrt{pq}}$.

Para comodidad de los lectores más adelante se expone la tabla 2 de la función de distribución normal $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$.

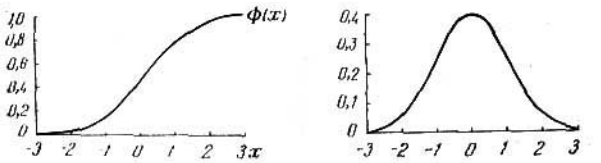

Fig. 14. Vista general de la función de distribución normal de probabilidades $\Phi(x)$ con los parámetros $(0, 1)$ y de la densidad de probabilidad respectiva.

2. Proceso del movimiento browniano. Distribución de probabilidades del máximo y momento de su consecución. Imaginémonos a una partícula suspendida en un líquido homogéneo. Ella sufre unos choques caóticos con las moléculas del líquido en resultado de lo cual se encuentra en un movimiento continuamente desordenado.

De modelo análogo discreto de tal proceso puede servir el siguiente modelo de la fluctuación aleatoria. La partícula cambia su posición sólo en momentos discretos de tiempo, múltiplos de Δt. El cambio de posición ocurre de tal forma que encontrándose en el punto x, la partícula independientemente del comportamiento anterior pase con las mismas probabilidades a uno de los puntos vecinos $x + \Delta x$ ó $x - \Delta x$, siendo una misma la desviación Δx para todos los puntos x (se trata sólo de una de las coordenadas de la partícula, en el caso de la fluctuación aleatoria unidimensional). En el límite, cuando de un modo determinado $\Delta t \to 0$, $\Delta x \to 0$ se obtiene la fluctuación aleatoria c o n t ï n u a, para el proceso físico del movimiento browniano[1]).

[1]) Véase, por ejemplo, el libro de A. Ya. Jinchin «Leyes asintóticas de la teoría de probabilidades» ONTI. 1936 (cap. 3).

Tabla 2

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,0	0,500 000	1,5	0,933 193	3,0	0,998 650
0,1	0,539 828	1,6	0,945 201	3,1	0,999 032
0,2	0,579 260	1,7	0,955 435	3,2	0,999 313
0,3	0,617 911	1,8	0,964 070	3,3	0,999 517
0,4	0,655 422	1,9	0,971 283	3,4	0,999 663
0,5	0,691 462	2,0	0,977 250	3,5	0,999 767
0,6	0,725 747	2,1	0,982 136	3,6	0,999 841
0,7	0,758 036	2,2	0,986 097	3,7	0,999 892
0,8	0,788 145	2,3	0,989 276	3,8	0,999 928
0,9	0,815 940	2,4	0,991 802	3,9	0,999 952
1,0	0,841 345	2,5	0,993 790	4,0	0,999 968
1,1	0,864 334	2,6	0,995 339	4,1	0,999 979
1,2	0,884 930	2,7	0,996 533	4,2	0,999 987
1,3	0,903 200	2,8	0,997 445	4,3	0,999 991
1,4	0,919 243	2,9	0,998 134	4,4	0,999 995
				4,5	0,999 997

Designemos por $\xi(t)$ la posición de la partícula browniana en el momento de tiempo t. Supongamos que en el momento inicial de tiempo $t = 0$ la partícula se encuentra en el punto $x = 0$.

Durante la fluctuación discreta en el tiempo t ella realiza $n = \frac{t}{\Delta t}$ pasos, entre los cuales se realiza un número aleatorio determinado de pasos en dirección positiva. Si designamos por S_n al número de pasos en dirección positiva, entonces el desplazamiento total en dirección positiva será $S_n \Delta x$ y en dirección negativa, $(n - S_n)\,\Delta x$. De tal modo, el desplazamiento total $\xi(t)$ en el tiempo $t = n\Delta t$ está ligado con el número S_n por la siguiente igualdad:

$$\xi(t) = [S_n\,\Delta x - (n - S_n)\,\Delta x] = (2S_n - n)\,\Delta x.$$

Si consideramos que $\xi(0) = 0$, entonces

$$\xi(s + t) = [\xi(s) - \xi(0)] + [\xi(t + s) - \xi(s)]$$

para cualesquiera s, $t \geqslant 0$. Evidentemente, en el modelo de la fluctuación aleatoria descrito, las magnitudes $\xi(s) - \xi(0)$ y $\xi(t + s) - \xi(s)$ son i n d e p e n d i e n t e s, siendo la distribución de probabilidades del incremento $\xi(t + s) - \xi(s)$ exactamente igual que la del incremento $\xi(t) - \xi(0)$. Por eso, para la dispersión $\mathbf{D}\xi(t + s)$

con cualesquiera s, $t \geqslant 0$, tiene lugar la igualdad

$$\mathbf{D}\xi(t+s) = \mathbf{D}[\xi(s) - \xi(0)] + \mathbf{D}[\xi(t+s) - \xi(s)] = \\ = \mathbf{D}\xi(s) + \mathbf{D}\xi(t).$$

Se ve que la dispersión $\mathbf{D}\xi(t)$ como función de t, varía linealmente con el crecimiento de t y de tal modo

$$\mathbf{D}\xi(t) = \sigma^2 \cdot t, \quad 0 \leqslant t \leqslant \infty,$$

donde σ^2 es una determinada constante llamada *coeficiente de difusión*. Por otra parte, se puede calcular, fácilmente, que la dispersión del desplazamiento durante el tiempo t (o de otra forma, durante n pasos, $n = \frac{t}{\Delta t}$) es $\mathbf{D}\xi(t) = (\Delta x)^2 \frac{t}{\Delta t}$. Como resultado obtenemos la siguiente relación entre Δx y Δt:

$$\frac{(\Delta x)^2}{\Delta t} = \sigma^2 .$$

Los pasos realizados por la partícula no dependen uno de otro, y los podemos considerar como pruebas de Bernoulli, con la probabilidad del «éxito» $p = 1/2$, que la tomamos como un paso en dirección positiva. Entonces S_n, que es el número de pasos en dirección positiva, será igual al número de «éxitos» en las n pruebas de Bernoulli. Con esto, la posición de la partícula en el momento t estará ligada a la magnitud normalizada $S_n^* = \frac{1}{\sqrt{n}}(2S_n - n)$ del siguiente modo:

$$\xi(t) = S_n^* \sqrt{n}\,\Delta x = S_n^* \sqrt{t}\,\frac{\Delta x}{\sqrt{\Delta t}} = S_n^* \sigma \sqrt{t}.$$

Utilizando el teorema de Moivre—Laplace (véase la relación (3.6)), obtenemos que la distribución de probabilidades de la magnitud aleatoria $\xi(t)$ en el límite se da por la fórmula

$$\mathbf{P}\left\{x' \leqslant \frac{\xi(t)}{\sigma\sqrt{t}} \leqslant x''\right\} = \lim_{\Delta t \to 0} \mathbf{P}\{x' \leqslant S_n^* \leqslant x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2}\,dx. \quad (3.9)$$

La fórmula (3.9) da no sólo la distribución de probabilidades de la magnitud $\xi(t)$, que es la posición de la partícula browniana en el momento t, sino también la de cualquier desplazamiento de la partícula que fluctúa aleatoriamente en el tiempo t, ya que debido a la homogeneidad del proceso examinado, para cualquier valor de $\xi(s)$, el incremento $\xi(t+s) - \xi(s)$ tiene la misma distribución de probabilidades que el incremento $\xi(t) - \xi(0) = \xi(t)$, a saber:

$$\mathbf{P}\left\{x' \leqslant \frac{\xi(t+s) - \xi(s)}{\sigma\sqrt{t}} \leqslant x''\right\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2}\,dx.$$

En la fig. 15[1]) están representadas algunas trayectorias experimentales del movimiento browniano.

El modelado del movimiento browniano se puede realizar con ayuda de la llamada «*tabla de Galton*». Este dispositivo representa en sí una tabla lisa con clavitos dispuestos simétricamente; se lanza desde arriba una bolita («partícula browniana») que en su movimiento hacia abajo tropieza con los clavitos («moléculas») y se desplaza cada

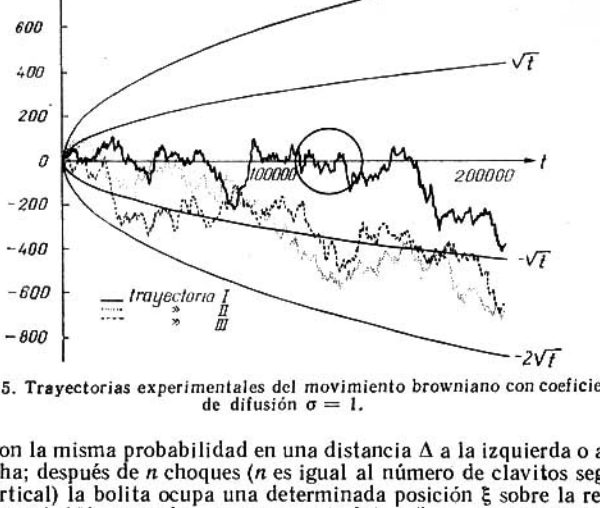

Fig. 15. Trayectorias experimentales del movimiento browniano con coeficiente de difusión $\sigma = 1$.

vez con la misma probabilidad en una distancia Δ a la izquierda o a la derecha; después de n choques (n es igual al número de clavitos según la vertical) la bolita ocupa una determinada posición ξ sobre la recta horizontal OX, cayendo en el intervalo $[x', x'']$ con la probabilidad

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} \approx \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2}\, dx$$

(consideramos que en la escala elegida $\Delta \sqrt{n} = 1$); de hecho la bolita cae en una de las celdas que dividen al eje OX en pequeños intervalos $[x_k, x_{k+1}]$ de igual longitud Δx.

Si tomamos un número N de bolitas, suficientemente grande, entonces de acuerdo a la ley de los grandes números, el número N_k

[1]) Véase H. O. A. Wold. (Editor), Bibliography on time series and stochastic processes, Oliver and Boyd, Edinburgh and London, 1965, (págs. 10-11).

de aquellas que caen en la celda k $[x_k, x_{k+1}]$, deberá ser tal que $\frac{N_k}{N} \approx$

$$\approx \mathbf{P}\{x_k \leqslant \xi \leqslant x_{k+1}\} \approx \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} \Delta x.$$

En el modelo discreto, para cualquier valor fijado $\xi(s) = a$, el movimiento de la partícula browniana a partir del momento s (cuando ella se encuentra en el punto a) no depende de su comportamiento anterior a este momento. Suponiendo que esta propiedad también se conserva en el proceso continuo límite del movimiento browniano, hallamos la distribución de probabilidades de la magnitud aleatoria τ_a, o sea, *el momento en que por primera vez la partícula browniana alcanza el punto* $x = a$ (considerando, como anteriormente, que la partícula se encuentra en el momento inicial $t = 0$ en el punto $x = 0$).

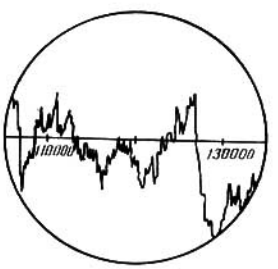

Fig. 15a. Parte de la fig. 15 aumentada 12 veces.

Está claro, que la partícula browniana durante el movimiento en dirección positiva, obedece a la misma ley, que durante el movimiento en dirección negativa (en el modelo discreto la partícula se mueve a cada paso a la derecha o a la izquierda con la misma probabilidad). Por eso, las magnitudes τ_a y τ_{-a} (momentos de alcanzar a los puntos a y $-a$), al salir del punto inicial $x = 0$, tienen la misma distribución de probabilidades. Consideramos que $a > 0$ y hallaremos la probabilidad $\mathbf{P}\{\tau_a \leqslant t\}$.

La partícula puede encontarse en el momento t más a la derecha del punto a, sólo con la condición de que, en determinado momento $\tau_a \leqslant t$, se encuentre en este punto (ya que durante el movimiento browniano continuo, la partícula no puede «saltar» por encima de a). Esto significa formalmente que el suceso $\{\xi(t) \geqslant a\}$ está contenido en el suceso $\{\tau_a \leqslant t\}$ y, por consiguiente,

$$\mathbf{P}\{\xi(t) \geqslant a \mid \tau_a \leqslant t\} = \frac{\mathbf{P}\{\xi(t) \geqslant a\}}{\mathbf{P}\{\tau_a \leqslant t\}}.$$

Evidentemente, la probabilidad condicional $\mathbf{P}\{\xi(t) \geqslant a \mid \tau_a \leqslant t\}$, a condición de que la partícula se encuentre en a en el momento τ_a ($\tau_a \leqslant t$), coincide con la probabilidad de que ella se encuentre más a la derecha de a en el momento t después de salir de a. Pero, de las consideraciones de simetría expuestas anteriormente, se deduce que la probabilidad de encontrarse en el momento t, más a la derecha del punto de partida a, es la misma que la probabilidad de encontrarse en este momento, más a la izquierda de a, e igual a 1/2. De tal forma,

$\mathbf{P}\{\xi(t) \geqslant a \mid \tau_a \leqslant t\} = 1/2$ y considerando, para sencillez, al coeficiente de difusión σ^2 igual a 1, de la igualdad obtenida antes y de la fórmula (3.9) obtenemos la función de distribución de la magnitud τ_a:

$$F_{\tau_a}(t) = \mathbf{P}\{\tau_a \leqslant t\} = 2\mathbf{P}\{\xi(t) \geqslant a\} = \sqrt{\frac{2}{\pi}} \int\limits_{at^{-1/2}}^{\infty} e^{-x^2/2}\, dx, \quad t > 0.$$

Diferenciando por t la función de distribución, hallamos la densidad de probabilidad correspondiente:

$$p_{\tau_a}(t) = \frac{a}{\sqrt{2\pi}}\, t^{-3/2} e^{-a^2/2t} \quad 0 \leqslant t < \infty \tag{3.10}$$

($p_{\tau_a}(t) = 0$ para $t < 0$).

Es interesante observar que para cualquier punto a la magnitud τ_a es finita con la probabilidad 1:

$$\mathbf{P}\{\tau_a < \infty\} = \lim_{t \to \infty} \mathbf{P}\{\tau_a \leqslant t\} = \sqrt{\frac{2}{\pi}} \int\limits_{0}^{\infty} e^{-x^2/2}\, dx = 1,$$

es decir, la partícula browniana cae más tarde o más pronto en cualquier punto a (en un determinado momento aleatorio de tiempo $\tau_a < \infty$).

Conociendo la distribución de la magnitud τ_x, o sea, el momento de alcanzar el punto x, se puede hallar también, inmediatamente, la distribución de probabilidades de la magnitud de *máximo desplazamiento* de la partícula browniana en el tiempo fijado t. Evidentemente,

$$\mathbf{P}\{\max_{0 \leqslant s \leqslant t} \xi(s) \geqslant x\} = \mathbf{P}\{\tau_x \leqslant t\} = \sqrt{\frac{2}{\pi}} \int\limits_{xt^{-1/2}}^{\infty} e^{-u^2/2}\, du =$$

$$= \sqrt{\frac{2}{\pi t}} \int\limits_{x}^{\infty} e^{-u^2/2t}\, du$$

y, por consiguiente, la función de distribución de la magnitud $\xi = \max\limits_{0 \leqslant s \leqslant t} \xi(s)$ será

$$F_{\xi}(x) = 1 - \mathbf{P}\{\max_{0 \leqslant s \leqslant t} \xi(s) \geqslant x\} = \sqrt{\frac{2}{\pi t}} \int\limits_{0}^{x} e^{-u^2/2t}\, du$$

$\left(\text{recordemos que } \sqrt{\frac{2}{\pi t}} \int\limits_{0}^{\infty} e^{-u^2/2t}\, dt = \frac{2}{\sqrt{2\pi}} \int\limits_{0}^{\infty} e^{-u^2/2}\, du = 1\right)$. Se ve, que la magnitud $\xi = \max\limits_{0 \leqslant s \leqslant t} \xi(s)$ tiene la densidad de probabilidad

$$p_{\xi}(x) = \sqrt{\frac{2}{\pi t}}\, e^{-x^2/2t}, \quad 0 \leqslant x < \infty \tag{3.11}$$

($p_\xi(x) = 0$ para $x < 0$, ya que $\xi \geqslant \xi(0) = 0$). Esta es la llamada *ley normal doble* de distribución de probabilidades (como se ve fácilmente, $\mathbf{P}\{\xi \geqslant x\} = 2\mathbf{P}\{\xi(t) \geqslant x\}$).

Evidentemente, la magnitud $\min\limits_{0 \leqslant s \leqslant t} \xi(s)$ tiene una distribución de probabilidades análoga, y precisamente, su densidad de probabilidad será

$$p(x) = \sqrt{\frac{2}{\pi t}}\, e^{-x^2/2t}, \qquad -\infty < x \leqslant 0$$

($p(x) = 0$ para $x > 0$). Es interesante indicar que

$$\mathbf{P}\{\max_{0 \leqslant s \leqslant t} \xi(s) > 0\} = \mathbf{P}\{\min_{0 \leqslant s \leqslant t} \xi(s) < 0\} = 1$$

y por lo tanto, *al salir del punto* $x = 0$, *la partícula browniana, en cualquier tiempo* t, *tan pequeño como se quiera, estará tanto más a la derecha del punto inicial* $x = 0$ *como más a la izquierda de este punto.* Examinando la trayectoria de la partícula browniana en el tiempo (es decir, el gráfico de la función $\xi = \xi(t)$, $t \geqslant 0$), de aquí se puede deducir que esta trayectoria continua en un intervalo de tiempo $(0, t)$ tan pequeño como se quiera, corta infinitas veces al nivel $x = 0$, tomando un número infinito de veces valores tanto positivos como negativos (en otras palabras, la partícula browniana retorna un número infinito de veces al punto inicial $x = 0$).

Suponiendo que es continua, la trayectoria de la partícula browniana $\xi(u)$, $0 \leqslant u \leqslant t$, alcanza su máximo absoluto en un punto determinado τ, $0 \leqslant \tau \leqslant t$ (tendremos en consideración el primer punto del máximo, si son varios). Hallemos la distribución de la magnitud aleatoria τ.

Supongamos que se tiene la densidad de la distribución conjunta de probabilidades de las magnitudes aleatorias τ y $\xi = \xi(\tau)$ ($\xi = \max\limits_{0 \leqslant u \leqslant t} \xi(u)$). Mostremos que entonces, la densidad tiene la forma

$$p_{\tau, \xi}(s, x) = \frac{1}{\pi \sqrt{s(t-s)}} \frac{x}{s}\, e^{-x^2/2s}, \qquad (3.12)$$

$$0 < s < t, \quad 0 \leqslant x < \infty.$$

Para ello, examinemos al principio la distribución conjunta de probabilidades de las magnitudes aleatorias τ_a y ξ, donde τ_a, como anteriormente, significa el momento en que la partícula browniana alcanza por primera vez el punto $a > 0$.

Después de caer la partícula browniana en el punto a su comportamiento ulterior se somete a las mismas leyes, como si este punto fuese el inicial desde el propio comienzo. Por eso la magnitud $\xi = \max\limits_{0 \leqslant u \leqslant t} \xi(u)$, que para la condición $\tau_a = s$, $0 < s \leqslant t$, coincide con la magnitud $\max\limits_{s \leqslant u \leqslant t} \xi(u)$, tiene, para la condición indicada,

la misma distribución de probabilidades que la magnitud $a + \max_{0 \leqslant u \leqslant t-s} \xi(u)$, y de acuerdo con la fórmula establecida anteriormente (3.11), tiene la densidad convencional de distribución

$$p_\xi(x \mid s) = \sqrt{\frac{2}{\pi(t-s)}}\, e^{-(x-a)^2/2(t-s)},$$
$$a \leqslant x < \infty.$$

De aquí se deduce, que la densidad $p_{\tau_a, \xi}(s, x)$ de la distribución conjunta de probabilidades de las magnitudes τ_a, ξ para $0 < s < t$, $a \leqslant x < \infty$ tiene la forma (véase (3.17), cap. I):

$$p_{\tau_a, \xi}(s, x) = p_{\tau_a}(s)\, p_\xi(x \mid s) = \frac{1}{\pi\sqrt{s(t-s)}} \frac{a}{s} e^{-a^2/2s} e^{-(x-a)^2/2(t-s)}.$$

Por otra parte, para la condición $\xi = \max_{0 \leqslant u \leqslant t} \xi(u) = a$ el punto τ del máximo coincide con el momento τ_a, y por consiguiente, para la condición indicada, las magnitudes τ, τ_a tienen la misma distribución de probabilidades. De aquí se desprende que la densidad de probabilidad de las magnitudes (τ, ξ) en el punto $\tau = s$, $\xi = a$, coincide con la densidad de probabilidad de las magnitudes (τ_a, ξ) en el mismo punto (s, a) ya que

$$p_{\tau, \xi}(s, a) = p_\tau(s \mid a)\, p_\xi(a) = p_{\tau_a}(s \mid a)\, p_\xi(a) = p_{\tau_a, \xi}(s, a),$$

donde $p_\tau(s \mid x)$ y $p_{\tau_a}(s \mid x)$ representan a las densidades convencionales de las distribuciones de probabilidades de las magnitudes τ y τ_a, para la condición $\xi = x$. De la fórmula hallada anteriormente para la densidad $p_{\tau_a, \xi}(x, s)$, cuando $x = a$, obtenemos

$$p_{\tau, \xi}(s, a) = \frac{1}{\pi\sqrt{s(t-s)}} \frac{a}{s} e^{-a^2/2s}, \quad 0 < s < t, \quad 0 < a < \infty,$$

lo que nos da la fórmula indicada anteriormente (3.12). La densidad de una magnitud τ tomada por separado, o sea del punto máximo de la trayectoria browniana $\xi(s)$ en el intervalo de tiempo $0 \leqslant s \leqslant t$ será

$$p_\tau(s) = \int_0^\infty p_{\tau, \xi}(s, x)\, dx = \frac{1}{\pi\sqrt{s(t-s)}} \int_0^\infty \frac{x}{s} e^{-x^2/2s}\, dx =$$
$$= \frac{1}{\pi\sqrt{s(t-s)}}, \quad 0 < s < t.$$

La función de distribución de probabilidades correspondiente tiene la forma

$$F_\tau(s) = \int_0^s \frac{du}{\pi\sqrt{u(t-u)}} = \frac{2}{\pi} \operatorname{arco\,sen} \sqrt{\frac{s}{t}}, \quad 0 \leqslant s \leqslant t.$$

Esta ley de distribución de probabilidades lleva el nombre de *ley del arco seno*. La misma distribución de probabilidades tiene también, evidentemente, el punto del mínimo de la trayectoria $\xi(s)$, $0 \leqslant s \leqslant t$.

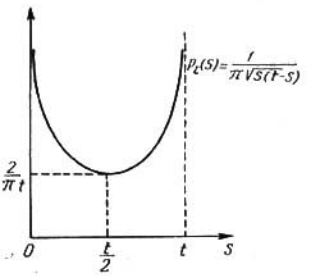

Fig. 16.

Se ve (fig. 16) que el comportamiento más probable de la partícula browniana es aquel, para el cual el punto extremal de su trayectoria está situado en la proximidad de los extremos del intervalo examinado $[0, t]$.

§ 4. DISTRIBUCION NORMAL DE PROBABILIDADES Y DISTRIBUCIONES LIGADAS A ELLA

1. Distribución normal multidimensional. Examinemos *la distribución normal* de probabilidades con la densidad

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \quad -\infty < x < \infty. \tag{4.1}$$

La densidad de probabilidad $p_0(x)$ es una función par, y por eso

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2}\, dx = 0$$

(es decir, el valor medio es igual a cero). Para calcular la dispersión $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx$, recurrimos a la "integral de probabilidad"

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx = 1.$$

Pongamos $x\sqrt{u}$ en lugar de x, para todos los valores $u>0$ tenemos

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ux^2/2}\,dx \equiv \frac{1}{\sqrt{u}}.$$

Diferenciando por el parámetro u en el punto $u=1$ obtenemos que la dispersión es

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,dx = 1.$$

Sea ξ_0 la magnitud aleatoria, que tiene la distribución normal de probabilidades (4.1); como acabamos de mostrar

$$M\xi_0 = 0, \quad D\xi_0 = 1.$$

Examinemos la magnitud aleatoria $\xi = \sigma\xi_0 + a$, $\sigma>0$. Está claro que

$$M\xi = a, \quad D\xi = \sigma^2, \tag{4.2}$$

y la densidad de la distribución de probabilidades de la magnitud ξ será

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-a)^2/2\sigma^2}, \quad -\infty < x < \infty. \tag{4.3}$$

La distribución de probabilidades con tal densidad se llama *distribución normal* (o *de Gauss*); se determina con dos parámetros (a, σ), *donde el parámetro a es el valor medio*:

$$a = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} x e^{-(x-a)^2/2\sigma^2}\,dx, \tag{4.4}$$

y *el parámetro* σ, así llamada *desviación standard* (normalizada), es la raíz positiva de la dispersión

$$\sigma^2 = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} (x-a)^2 e^{-(x-a)^2/2\sigma^2}\,dx. \tag{4.5}$$

Para los valores de los parámetros $a=0$ y $\sigma=1$, la distribución (4.3) corrientemente se llama distribución de Gauss standard.

Examinemos n magnitudes aleatorias independientes $\xi_{01}, \ldots, \xi_{0n}$, que tienen una misma distribución de Gauss (standard). Su densidad conjunta de probabilidades es

$$p_0(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{k=1}^{n} x_k^2}, \quad -\infty < x_1, \ldots, x_n < \infty. \tag{4.6}$$

Sea

$$\xi_i = \sum_{k=1}^{n} \sigma_{ik}\xi_{0k} + a_i, \quad i = 1, \ldots, n, \tag{4.7}$$

una transformación lineal no degenerada de las magnitudes ξ_{0k}, $k = 1, \ldots, n$. Evidentemente,

$$\mathbf{M}\xi_i = a_i, \quad \mathbf{M}(\xi_i - a_i)(\xi_j - a_j) = \sum_{k=1}^{n} \sigma_{ik}\sigma_{jk}, \quad i, j = 1, \ldots, n, \tag{4.8}$$

ya que las magnitudes independientes ξ_{0k}, $k = 1, \ldots, n$, con los valores medios nulos son no correlacionados:

$\mathbf{M}\xi_{0k} = 0$, $\mathbf{M}(\xi_{0k}\xi_{0l}) = \mathbf{M}\xi_{0k} \cdot \mathbf{M}\xi_{0l} = 0$ para $k \neq l$, $k, l = 1, \ldots, n$.

Recordemos que la matriz $R = \{R_{kj}\}$ con los elementos

$$R_{kj} = \mathbf{M}(\xi_i - a_i)(\xi_j - a_j), \quad i, j = 1, \ldots, n,$$

se llama *matriz de correlación* de las magnitudes aleatorias $\xi_1, \ldots, \xi_n$ $\left(r_{ij} = \frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}}\right.$ es el coeficiente de correlación entre las magnitudes ξ_i, $\left.\xi_j\right)$. Las fórmulas (4.8) muestran que la matriz de correlación R en el caso examinado tiene la forma

$$R = \sigma\sigma^*, \tag{4.9}$$

donde $\sigma = \{\sigma_{ij}\}$ es la matriz de la transformación lineal (4.7) y σ^* es la matriz conjugada con σ (con los elementos $\sigma^*_{ij} = \sigma_{ji}$, $i, j = 1, \ldots, n$). Los determinantes $|R|$ y $|\sigma|$ de las matrices indicadas están ligados por la igualdad $|R| = |\sigma|^2$; siendo $|\sigma|$ el jacobiano de la transformación (4.7) y, por consiguiente, la densidad conjunta de distribución de las magnitudes aleatorias $\xi_1, \ldots, \xi_n$ puede ser descrita por la fórmula

$$p(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}|R|} \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{n} b_{ij}(x_i - a_i)(x_j - a_j)\right\}, \tag{4.10}$$

donde $\sum_{i,j=1}^{n} b_{ij}(x_i - a_i)(x_j - a_j)$ es aquella forma cuadrática en la que se transforma la suma de los cuadrados $\sum_{k=1}^{n} x_k^2$ durante la transformación lineal inversa a (4.7). Se ve fácilmente, que *la matriz $\{b_{ij}\}$ de esta forma cuadrática es la inversa de la matriz de correlación $\{R_{ij}\}$*, ya que la suma de los cuadrados $\sum_{k=1}^{n} x_k^2$ se transforma

en la forma cuadrática

$$\sum_{k=1}^{n}\Big[\sum_{j=1}^{n} c_{kj}(x_j-a_j)\Big]^2=\sum_{i,j=1}^{n}\Big(\sum_{k=1}^{n} c_{ki}c_{kj}\Big)(x_i-a_i)(x_j-a_j),$$

donde $\{c_{kj}\}$ es la matriz inversa a $\{\sigma_{ij}\}$, de modo que

$$\{b_{ij}\}=\{c_{ik}\}^*\{c_{kj}\}=[\{\sigma_{ik}\}\{\sigma_{kj}\}^*]^{-1}=\{R_{ij}\}^{-1}.$$

Como ya se indicó anteriormente, *la matriz de correlación* $\{R_{ij}\}$ *está determinada positivamente*: para cualesquiera valores reales $c_1, \ldots, c_n$

$$\sum_{i,j=1}^{n} R_{ij}c_ic_j=\mathbf{M}\Big[\sum_{j=1}^{n} c_i(\xi_i-a_i)\Big]^2\geqslant 0;$$

la matriz inversa $\{b_{ij}\}=R^{-1}$ posee esta misma propiedad. Para cualquier matriz no degenerada, determinada positivamente, $\{b_{ij}\}$,

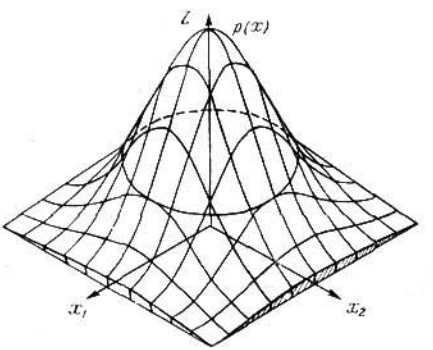

Fig. 17. Vista general de la distribución normal bidimensional (de la densidad de probabilidades $p(x)$); en la figura están indicadas las líneas de intersección de la superficie $z = p(x)$ con los distintos planos.

la fórmula (4.10), donde $|R|^{-1}$ significa el determinante de la matriz $\{b_{ij}\}$ y $a_1, \ldots, a_n$ son constantes arbitrarias, da una determinada densidad de probabilidad (esta forma tiene, precisamente, la densidad de probabilidad de las magnitudes $\xi_1, \ldots, \xi_n$, ligadas con las magnitudes de Gauss standards $\xi_{01}, \ldots, \xi_{0n}$ a través de la transformación lineal (4.7) con matriz $\{\sigma_{kj}\}$ que es la raíz cuadrada de la matriz $R=\{b_{ij}\}^{-1}$).

La distribución de probabilidades con la densidad (4.10), igualmente que las magnitudes aleatorias $\xi_1, \ldots, \xi_n$ con tal distribución de probabilidades, se llama *distribución normal* (o *de Gauss*) (fig. 17).

Evidentemente, durante cualquier transformación lineal no degenerada, las magnitudes de Gauss $\xi_1, \ldots, \xi_n$ se transforman en las magnitudes de Gauss $\eta_1, \ldots, \eta_n$: si

$$\eta_k = \sum_{j=1}^{n} c_{kj}\xi_j, \quad k = 1, \ldots, n,$$

entonces la densidad de la probabilidad de la magnitud $\eta_1, \ldots, \eta_n$ se da por la fórmula del tipo (4.10), a saber:

$$p_{\eta_1, \ldots, \eta_n}(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{n/2} |\vec{R}|^{1/2}} \times$$

$$\times \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{n} \vec{b}_{ij}(y_i - \vec{a}_i)(y_j - \vec{a}_j)\right\}, \quad (4.11)$$

donde

$$\vec{a}_i = \sum_{j=1}^{n} c_{ij}a_j, \quad i = 1, \ldots, n,$$

$$\{\vec{b}_{ij}\} = \{c_{ki}\}^{-1}\{b_{ki}\}\{c_{lj}\}^{-1}$$

y $|\vec{R}|$ es el determinante de la matriz $\vec{R} = \{\vec{b}_{ij}\}^{-1}$ (véase la fórmula (1.7) para la transformación de la densidad de probabilidades, cap. I).

Examinemos las magnitudes aleatorias de Gauss $\xi_0, \xi_1, \ldots, \xi_n$ y hallemos la distribución condicional de probabilidades de la magnitud ξ_0 cuando están fijados $\xi_1, \xi_2, \ldots, \xi_n$.

Consideremos la magnitud $\hat{\xi}_0 = a_0 + \sum_{k=1}^{n} \hat{c}_k(\xi_k - a_k)$, donde $a_k = \mathbf{M}\xi_k$, $k = 0, 1, \ldots, n$, y los coeficientes $\hat{c}_1, \ldots, \hat{c}_n$ satisfacen al sistema de ecuaciones lineales

$$\sum_{k=1}^{n} R_{kj}\hat{c}_k = R_{0j}, \quad j = 1, \ldots, n \quad (4.12)$$

($\hat{\xi}_0$ es la proyección de la magnitud ξ_0 en el espacio de todas las magnitudes del tipo $\sum_{k=1}^{n} c_k\xi_k$, véase (4.32) cap. I).

La transformación lineal no degenerada

$$\eta_0 = \hat{\xi}_0 - \hat{\xi}_0, \quad \eta_k = \xi_k - a_k, \quad k = 1, \ldots, n,$$

nos da las magnitudes de Gauss $\eta_0, \eta_1, \ldots, \eta_n$ tales que $\mathbf{M}\eta_0\eta_k = 0$ para todos los $k = 1, \ldots, n$; su matriz de correlación tiene la forma

$$\vec{R} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & R \end{pmatrix},$$

donde $\sigma^2 = M\eta_0^2$ y $R = \{R_{ij}\}$ es la matriz de correlación de las magnitudes $\eta_1, \ldots, \eta_n$. Se puede comprobar fácilmente que la matriz inversa a $\hat{R}$ será

$$\hat{b} = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & R^{-1} \end{pmatrix}$$

y, por consiguiente, la densidad de probabilidad de las magnitudes de Gauss $\eta_0, \eta_1, \ldots, \eta_n$ tiene la forma

$$p_{\eta_0, \eta_1, \ldots, \eta_n}(y, y_1, \ldots, y_n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} \times$$

$$\times \frac{1}{(2\pi)^{(n-1)/2} |R|^{1/2}} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{n} b_{ij} y_i y_j\right\}. \quad (4.13)$$

Como se observa con facilidad ésta descompone en el producto de la densidad de probabilidad

$$p_{\eta_0}(y) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p_{\eta_0, \eta_1, \ldots, \eta_n}(y, y_1, \ldots, y_n)\, dy_1 \ldots dy_n =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-y^2/2\sigma^2} \quad (4.14)$$

de la magnitud η_0 y la densidad de probabilidad

$$p_{\eta_1, \ldots, \eta_n}(y_1, \ldots, y_n) = \int_{-\infty}^{\infty} p_{\eta_0, \eta_1, \ldots, \eta_n}(y, y_1, \ldots, y_n)\, dy =$$

$$= \frac{1}{(2\pi)^{(n-1)/2} |R|^{1/2}} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{n} b_{ij} y_i y_j\right\} \quad (4.15)$$

de las magnitudes $\eta_1, \ldots, \eta_n$. Esto significa que la magnitud η_0 no depende de $\eta_1, \ldots, \eta_n$. Por esto la distribución condicional de probabilidades de esta magnitud (para cualesquiera valores fijados $\eta_1, \ldots, \eta_n$) tiene una misma densidad normal $p_{\eta_0}(y)$ como la indicada anteriormente.

A nosotros nos interesa la magnitud aleatoria $\xi_0 = \eta_0 + \hat{\xi}_0$, donde $\hat{\xi}_0 = a_0 + \sum_{k=1}^{n} \hat{c}_k \eta_k$ y $\eta_k = \xi_k - a_k$, $k = 1, \ldots, n$. Evidentemente, para los valores fijados $\xi_1 = x_1, \ldots, \xi_n = x_n$, la magnitud ξ_0 se diferencia de η_0 por el sumando constante $\hat{x}_0 = a_0 + \sum_{k=1}^{n} \hat{c}_k (x_k - a_k)$ y, por consiguiente, tiene la densidad condicional de probabilidad

$$p_{\xi_0}(x \mid x_1, \ldots, x_n) = p_{\eta_0}(x - \hat{x}_0) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\hat{x}_0)^2/2\sigma^2}, \quad -\infty < x < \infty. \quad (4.16)$$

Esta es la densidad normal de probabilidad que responde al valor medio

$$\hat{x}_0 = \mathbf{M}(\xi_0 \mid x_1, \ldots, x_n) = \int_{-\infty}^{\infty} x p_{\xi_0}(x \mid x_1, \ldots, x_n)\, dx =$$

$$= a_0 + \sum_{k=1}^{n} \hat{c}_k (x_k - a_k)$$ y a la dispersión

$$\sigma^2 = \mathbf{M}\left[\xi_0 - a_0 - \sum_{k=1}^{n} \hat{c}_k (\xi_k - a_k)\right]^2.$$

Así pues, hemos hallado la distribución condicional de la magnitud ξ_0 para los valores fijados $\xi_1, \ldots, \xi_n$. De paso hemos establecido efectivamente los siguientes hechos: *si las magnitudes de Gauss no están correlacionadas, entonces son independientes*; *si la magnitud multidimensional* $(\xi_0, \xi_1, \ldots, \xi_n)$ *tiene una distribución normal, entonces la magnitud multidimensional* $(\xi_1, \ldots, \xi_n)$ *de un número menor de dimensiones, tendrá una distribución de probabilidades del mismo tipo*, y en particular, las componentes $\xi_1, \ldots, \xi_n$ tomadas por separado tienen una distribución normal de probabilidades del tipo (4.3) compárese con (4.13)—(4.15).

Nos podemos imaginar fácilmente, que en las aplicaciones frecuentemente tratamos magnitudes aleatorias compuestas de un gran número de componentes independientes. Resulta que para unas condiciones amplias tales magnitudes tienen una distribución normal de probabilidades.

Este hecho fundamental se expresa matemáticamente en la forma de tal o cual (en dependencia de las condiciones) *teorema límite central* (que es un caso muy particular del teorema de Moivre—Laplace). De ejemplo, puede servir el teorema límite central para la suma $S_n = \sum_{k=1}^{n} \xi_k$ de sumandos independientes $\xi_1, \xi_2, \ldots$, que satisfacen *a la condición de Liapunov*:

$$\frac{1}{B_n^3} \sum_{k=1}^{n} \mathbf{M}\,|\xi_k - a_k|^3 \to 0 \text{ para } n \to \infty \qquad (4.17)$$

($a_k = \mathbf{M}\xi_k$, $k = 1, 2, \ldots$; $B_n = \sqrt{\mathbf{D}S_n}$); precisamente, *con la condición* (4.17) *para la distribución normal de probabilidades de la suma normalizada*

$$S_n^* = \frac{S_n - \mathbf{M}S_n}{\sqrt{\mathbf{D}S_n}}$$

tiene lugar la relación límite siguiente[1]):

$$\lim_{n\to\infty} \mathbf{P}\{x' \leqslant S_n^* \leqslant x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2}\, dx. \tag{4.18}$$

Se puede precisar que si las magnitudes aleatorias $\xi_1, \xi_2, \ldots$ tienen la misma distribución de probabilidades, entonces para la función de distribución $F_n(x) = \mathbf{P}\{S_n^* \leqslant x\}$ de la magnitud S_n^* es justa la siguiente valuación:

$$\sup_{-\infty<x<\infty} |F_n(x) - \Phi(x)| \leqslant C_0 \frac{\mu}{\sigma^3} \frac{1}{\sqrt{n}}, \tag{4.19}$$

donde $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du$ es la función de distribución normal y C_0 es una constante[2])

$$\left(\mu = \mathbf{M}|\xi_k - a|^3, \quad a = \mathbf{M}\xi_k, \quad \sigma = \sqrt{\mathbf{D}\xi_k} \qquad k = 1, 2, \ldots\right).$$

2. Valuación de los parámetros de la distribución normal. La distribución χ^2 y la distribución de Student. La densidad normal de probabilidad se determina por dos parámetros: por el valor medio de a y la desviación standard (varianza o dispersión) σ:

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-a)^2/2\sigma^2}, \quad -\infty < x < \infty$$

(véase (4.3) — (4.5)).

Supongamos que tratamos la magnitud aleatoria distribuida según la ley normal, cuyos parámetros (a, σ) son desconocidos. Se pueden obtener las valuaciones estadísticas de estos parámetros. Precisamente, si $\xi_1, \ldots, \xi_n$ son magnitudes de Gauss independientes con una media a y la dispersión σ, entonces en calidad de valuación para la esperanza matemática desconocida a (según los valores de elección de que disponemos $\xi_1, \ldots, \xi_n$) se puede tomar la «media empírica»

$$\hat{a} = \frac{1}{n} \sum_{k=1}^{n} \xi_k \tag{4.20}$$

y para la dispersión σ^2, la magnitud

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (\xi_k - \hat{a})^2. \tag{4.21}$$

[1]) Véase más adelante el § 5.
[2]) Está calculado que $C_0 < 1$.

El valor medio empírico $\hat{a} = \frac{1}{n}\sum_{k=1}^{n} \xi_k$, siendo una combinación lineal de magnitudes de Gauss independientes $\xi_1, \ldots, \xi_n$, *tiene la distribución normal de probabilidades con los parámetros* a, $\frac{\sigma}{\sqrt{n}}$ $\left(a = \mathbf{M}\hat{a},\ \frac{\sigma^2}{n} = \mathbf{D}\hat{a}\right)$.

Para hallar la distribución de probabilidades de «la dispersión empírica» $\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} (\xi_k - \hat{a})^2$ nos dirigimos a la identidad

$$n\hat{\sigma}^2 = \sum_{j=1}^{n} \xi_j^2 - n\hat{a}^2 = \sum_{j=1}^{n} \xi_j^2 - \eta^2,$$

donde

$$\eta = \frac{\xi_1}{\sqrt{n}} + \ldots + \frac{\xi_n}{\sqrt{n}}.$$

Utilizando cualquier transformación o r t o g o n a l de las variables $\xi_1, \ldots, \xi_n$

$$\eta_k = \sum_{j=1}^{n} c_{kj}\xi_j, \quad k = 1, \ldots, n, \tag{4.22}$$

en la cual $\eta_n = \frac{\xi_1}{\sqrt{n}} + \ldots + \frac{\xi_n}{\sqrt{n}}$ obtenemos que

$$n\hat{\sigma}^2 = \sum_{j=1}^{n} \xi_j^2 - \eta^2 = \sum_{k=1}^{n} \eta_k^2 - \eta_n^2 = \sum_{k=1}^{n-1} \eta_k^2,$$

ya que en la transformación ortogonal la suma de los cuadrados $\sum_{j=1}^{n} \xi_j^2$ se transforma en suma de los cuadrados $\sum_{k=1}^{n} \eta_k^2$.

Evidentemente, «la dispersión empírica»

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} \left(\xi_k - \frac{1}{n}\sum_{j=1}^{n} \xi_j\right)^2$$

no se cambia al pasar de las magnitudes $\xi_1, \ldots, \xi_n$ a las $\xi_1 - a, \ldots, \xi_n - a$, de modo que sin limitar la generalidad, se puede considerar el valor medio a igual a cero. Entonces, la densidad conjunta de probabilidad de las magnitudes de Gauss independientes $\xi_1, \ldots, \xi_n$ será

$$p(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n} x_j^2\right\},$$

y la densidad conjunta de probabilidad de las magnitudes $\eta_1, \ldots, \eta_n$, que se obtiene con la transformación ortogonal (4.22), es

$$p(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n} y_k^2\right\}$$

(véase la fórmula de transformación de la densidad (4.11)). Se ve, que $\eta_1, \ldots, \eta_n$ son magnitudes de Gauss independientes entre sí y, por consiguiente, $n\hat{\sigma}^2 = \sum_{k=1}^{n-1} \eta_k^2$ es la suma de los cuadrados de las $n-1$ magnitudes de Gauss independientes entre sí η_k con los parámetros $(0, \sigma)$. También vemos que *las valuaciones examinadas*

$$\hat{a} = \frac{1}{n}\eta_n \quad y \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n-1} \eta_k^2 \tag{4.23}$$

son magnitudes independientes.

La distribución de probabilidades de la magnitud aleatoria

$$\chi^2 = \sum_{k=1}^{n} \xi_{0k}^2, \tag{4.24}$$

donde $\xi_{01}, \ldots, \xi_{0n}$ son las magnitudes de Gauss independientes s t a n d a r d s, lleva el nombre de *ji-cuadrado distribución*; por n se designa el número de grados de libertad. Esta distribución tiene la densidad de la forma [1])

$$p(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2-1}e^{-x/2}, \quad 0 \leqslant x < \infty, \tag{4.25}$$

y es un caso particular de la distribución Γ (véase (2.17)) (fig. 18).

Señalemos que de la fórmula (4.25) se puede obtener, fácilmente, la expresión para la densidad de probabilidad de la magnitud aleatoria

$\chi = \sqrt{\sum_{k=1}^{n} \xi_{0k}^2}$, que es igual al módulo del vector $(\xi_{01}, \ldots, \xi_{0n})$:

$$p_\chi(x) = 2xp(x^2) = \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2}e^{-x^2/2}, \quad 0 \leqslant x < \infty; \tag{4.26}$$

para $n = 2$ esta es la tal llamada *distribución de Rayleigh*, para $n = 3$ es *la distribución de Maxwell*.

Así pues, para apreciar los valores $\hat{a}$, $\hat{\sigma}$ de los parámetros a, σ de la distribución normal (véase (4.23)) hemos establecido que $\hat{a}$

[1]) La fórmula (4.25) será deducida en el siguiente § 5, véase (5.21).

tiene una distribución normal de probabilidades con los parámetros $\left(a, \frac{\sigma}{\sqrt{n}}\right)$ *y* $n\frac{\hat{\sigma}^2}{\sigma^2}$ *tiene la* χ^2*-distribución con* $n-1$ *grados de libertad, siendo* $\hat{a}$ *y* $\hat{\sigma}$ *magnitudes aleatorias independientes.*

Si el parámetro σ es conocido, entonces para cualquier α tan pequeño como se quiera se puede indicar Δ de modo tal que el valor desco-

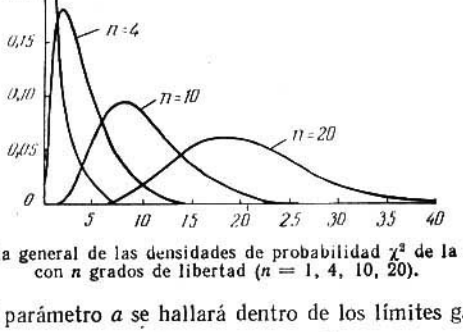

Fig. 18. Vista general de las densidades de probabilidad χ^2 de la distribución con n grados de libertad ($n = 1, 4, 10, 20$).

nocido del parámetro a se hallará dentro de los límites garantizados

$$\hat{a} - \frac{\sigma}{\sqrt{n}}\Delta \leqslant a \leqslant \hat{a} + \frac{\sigma}{\sqrt{n}}\Delta \qquad (4.27)$$

con una probabilidad no menor de $1-\alpha$, donde Δ se determina según α de la relación

$$\mathbf{P}\left\{\left|\frac{\hat{a}-a}{\sigma}\sqrt{n}\right| \leqslant \Delta\right\} = 2\Phi(\Delta) - 1 = 1 - \alpha$$

(Δ puede ser hallado en la tabla de la función normal de distribución $\Phi(x)$).

Si son desconocidos los dos parámetros (a, σ), entonces los límites garantizados para el valor medio de a, pueden ser establecidos sobre la base de la distribución que satisface a la relación

$$\tau = \sqrt{n-1}\,(\hat{a}-a)/\hat{\sigma} = \eta_n \Bigg/ \sqrt{\frac{1}{n-1}\sum_{k=1}^{n-1}\eta_k^2},$$

donde $\eta_1, \ldots, \eta_n$ son las magnitudes independientes de Gauss con los parámetros $(0, \sigma)$.

Está claro que la magnitud $\tau = \eta_n \Big/ \sqrt{\frac{1}{n-1}\sum_{k=1}^{n-1}\eta_k}$ no depende

de σ, y su distribución de probabilidades coincide con la distribución de la relación $\xi_{0n} \Big/ \sqrt{\frac{1}{n-1}\sum_{j=1}^{n-1}\xi_{0j}}$ para las magnitudes de Gauss "standards" independientes $\xi_{01}=\frac{\eta_1}{\sigma}$, ..., $\xi_{0n}=\frac{\eta_n}{\sigma}$. Esta distribución que se denomina *distribución de Student* (*o distribución t*)

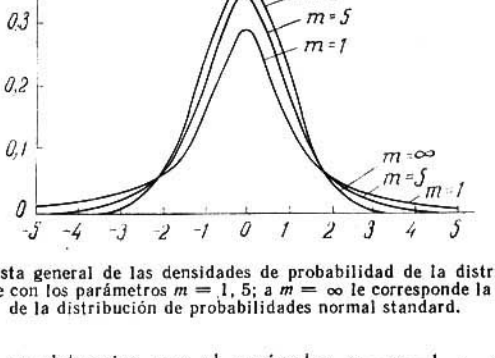

Fig. 19. Vista general de las densidades de probabilidad de la distribución *t* de Studente con los parámetros $m = 1, 5$; a $m = \infty$ le corresponde la densidad de la distribución de probabilidades normal standard.

(fig. 19), se determina por el parámetro $m = n-1$ y, como se indicará más adelante, tiene una densidad de la forma

$$p(x)=\frac{1}{\sqrt{\pi m}}\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}\left(1+\frac{x^2}{m}\right)^{-(m+1)/2}, \quad 0\leqslant x<\infty. \quad (4.28)$$

Igualmente que durante la determinación del «intervalo confidencial» (4.27), se puede hallar el valor correspondiente Δ para el cual

$$\mathbf{P}\left\{\left|\sqrt{n-1}\,\frac{\hat{a}-a}{\hat{\sigma}}\right|\leqslant\Delta\right\}=\mathbf{P}\{|\tau|\leqslant\Delta\}=1-\alpha,$$

y, por consiguiente, *con la probabilidad* $1-\alpha$ *se puede garantizar que*

$$\hat{a}-\frac{\hat{\sigma}}{\sqrt{n-1}}\Delta\leqslant a\leqslant\hat{a}+\frac{\hat{\sigma}}{\sqrt{n-1}}\Delta. \quad (4.29)$$

En conclusión deduzcamos la fórmula (4.28) para la densidad de la distribución t (la densidad de la probabilidad de la relación $\eta =$

$= \xi_1/\xi_2$ de las magnitudes de Gauss independientes ξ_1 con los parámetros (0, 1) y de la magnitud $\xi_2 = \sqrt{\frac{1}{m}\chi^2}$, donde la magnitud χ^2 tiene la χ^2-distribución con m grados de libertad).

De la fórmula (4.26) hallamos fácilmente que la densidad $p_2(x)$ de la magnitud ξ_2 es

$$p_2(x) = \frac{m^{m/2}}{2^{m/2-1}\Gamma\left(\frac{m}{2}\right)} x^{m-1} e^{-mx^2/2}, \quad x > 0.$$

Para la densidad de probabilidad de la relación $\eta = \xi_1/\xi_2$, teniendo en cuenta que $p_2(x) = 0$ para $x < 0$, tendremos la siguiente expresión:

$$p_\eta(y) = \int_0^\infty p_1(yx)\, p_2(x)\, x\, dx$$

(véase (3.14), cap. I). Poniendo aquí la expresión hallada anteriormente, para $p_2(x)$ y

$$p_1(yx) = \frac{1}{\sqrt{2\pi}} e^{-y^2x^2/2}, \quad -\infty < x < \infty,$$

obtenemos que

$$p_\eta(y) = \int_0^\infty \frac{m^{m/2}}{\sqrt{\pi}\, 2^{m/2}\, \Gamma\left(\frac{m}{2}\right)} x^{m-1} e^{-\frac{m}{2}x^2\left(1+\frac{y^2}{m}\right)} dx.$$

Haciendo en la integral el cambio de variable $u = \frac{m}{2}x^2 \times \left(1+\frac{y^2}{m}\right)$, llegamos a la expresión

$$p_\eta(y) = \frac{1}{\sqrt{\pi m}\, \Gamma\left(\frac{m}{2}\right)} \left(1+\frac{y^2}{m}\right)^{-(m+1)/2} \int_0^\infty u^{(m-1)/2} e^{-u}\, du,$$

en el cual

$$\int_0^\infty u^{(m-1)/2} e^{-u}\, du = \Gamma\left(\frac{m+1}{2}\right)$$

es el valor de la función Γ en el punto $(m+1)/2$. Hemos obtenido la fórmula (4.28);

§ 5. DISTRIBUCION DE PROBABILIDADES Y FUNCIONES CARACTERISTICAS

1. Funciones características y sus propiedades fundamentales. La función

$$\varphi(t) = Me^{it\xi}, \quad -\infty < t < \infty, \tag{5.1}$$

cuyo valor en el punto t coincide con la esperanza matemática $Me^{it\xi}$, donde ξ es la magnitud aleatoria real, se denomina *función característica* de esta magnitud ξ (o de la correspondiente distribución de probabilidades con la función de distribución $F(x) = \mathbf{P}\{\xi < x\}$).

Supongamos que la distribución de probabilidades está concentrada en los puntos enteros $k = 0, \pm 1, \pm 2, \ldots$ (es decir, $\sum_{k=-\infty}^{\infty} P(k) = 1$, donde $P(k) = \mathbf{P}\{\xi = k\}$, $k = 0, \pm 1, \pm 2, \ldots$), entonces la función característica

$$\varphi(t) = \sum_{k=-\infty}^{\infty} e^{itk} P(k) \tag{5.2}$$

es periódica (con período 2π) y la fórmula (5.2) da su descomposición en serie de Fourier con los coeficientes

$$P(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi(t)\, dt, \quad k = 0, \pm 1, \pm 2, \ldots \tag{5.3}$$

Supongamos que la distribución de probabilidades tiene la densidad $p(x)$, entonces la función característica es

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} p(x)\, dx \tag{5.4}$$

y representa en sí la integral de Fourier de la función $p(x)$, $-\infty < x < \infty$, y para la función absolutamente integrable $\varphi(t) = \left(\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty \right)$ la densidad $p(x)$ puede ser obtenida con ayuda de la transformación inversa de Fourier:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \varphi(t)\, dt, \quad -\infty < x < \infty. \tag{5.5}$$

Exponemos *la fórmula de inversión* general[1]), que liga la función característica $\varphi(t)$ y la función de distribución correspondiente $F(x)$:

$$F(x'')-F(x')=\lim_{\sigma\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ix''t}-e^{-ix't}}{-it}e^{-\sigma^2t^2/2}\varphi(t)\,dt \qquad (5.6)$$

para cualesquiera puntos x' y x'' en los cuales la función de distribución $F(x)$ es continua (es decir, $\mathbf{P}\{\xi=x'\}=\mathbf{P}\{\xi=x''\}=0$).

Señalemos aquí que se tiene no más de n valores diferentes de x, para los cuales $\mathbf{P}\{\xi=x\}\geqslant\frac{1}{n}$, y esto significa, que se tiene un número finito y numerable de valores de x, para los cuales $\mathbf{P}\{\xi=x\}>0$; es evidente que para cualquier x', x'' se hallarán sucesiones de puntos $x_n\to x'-0$, $x_n''\to x''+0$ que satisfacen a la condición $\mathbf{P}\{\xi=x_n'\}=\mathbf{P}\{\xi=x_n''\}=0$, y debido a la continuidad de la probabilidad,

$$\lim_{n\to\infty}\mathbf{P}\{x_n'\leqslant\xi\leqslant x_n''\}=\mathbf{P}\{x'\leqslant\xi\leqslant x''\}=$$
$$=F(x'')-F(x'-0), \qquad (5.7)$$

lo que da, junto con la fórmula de inversión, el siguiente resultado: *la distribución de probabilidades se determina univocamente por la función característica.*

Examinemos el desarrollo

$$e^{it\xi}=\sum_{k=0}^{n}i^k\frac{\xi^k}{k!}t^k+\theta\frac{\xi^{n+1}}{(n+1)!}t^{n+1},$$

donde θ es una magnitud limitada, $|\theta|\leqslant 1$. Si se tienen los momentos finitos $\mu_k=\mathbf{M}|\xi|^k$, $k=1,\ldots,n+1$, entonces, evidentemente,

$$\mathbf{M}e^{it\xi}=\sum_{k=0}^{n}i^k\frac{\mathbf{M}\xi^k}{k!}t^k+\frac{\mathbf{M}(\theta\xi^{n+1})}{(n+1)!}t^{n+1},$$

y de tal modo, la función característica $\varphi(t)$ admite el desarrollo

$$\varphi(t)=\sum_{k=0}^{n}i^k\frac{\mu_k}{k!}t^k+\frac{R_n}{(n+1)!}t^{n+1},$$
$$-\infty<t<\infty, \qquad (5.8)$$

donde $|R_n|\leqslant\mathbf{M}|\xi|^{n+1}$ y $\mu_0=\varphi(0)=1$,
$\mu_k=i^{-k}\varphi^k(0)$, $k=1,\ldots,n$.

La fórmula (5.8) permite calcular los momentos $\mu_k=\mathbf{M}\xi^k$ según las derivadas de la función característica de la magnitud aleatoria ξ.

[1]) Semejantes fórmulas de inversión serán examinadas más adelante, véase la pág. 130.

Examinemos algunos ejemplos.

Ejemplo (distribución uniforme). Para densidad de probabilidad

$$p(x)=\begin{cases}\dfrac{1}{b-a} & \text{para } a\leqslant x\leqslant b\\ 0 & \text{para } x<a,\ x>b\end{cases}$$

la función característica es

$$\varphi(t)=\frac{1}{b-a}\int_a^b e^{itx}\,dx=\frac{e^{ibt}-e^{iat}}{(b-a)\,it}.$$

$$-\infty<t<\infty.$$

Ejemplo (distribución de Poisson). Hallemos la función característica de la función de distribución de Poisson

$$P(k)=\frac{a^k}{k!}e^{-a},\ k=0,\ 1,\ \ldots$$

Tenemos

$$\varphi(t)=\sum_{k=0}^{\infty}e^{itk}P(k)=e^{-a}\sum_{k=0}^{\infty}\frac{(ae^{it})^k}{k!}=$$

$$=e^{-a}e^{ae^{it}}=e^{a(e^{it}-1)},\ -\infty<t<\infty. \tag{5.9}$$

Ejemplo (distribución normal). Hallemos la función característica de la distribución normal con la densidad

$$p(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},\ -\infty<x<\infty.$$

Para el parámetro real z tenemos

$$\int_{-\infty}^{\infty}e^{zx}p(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{zx-x^2/2}\,dx=e^{z^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-(x+z)^2/2}\,dx=$$

$$=e^{z^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}\,dx=e^{z^2/2},$$

de modo que la integral $\int_{-\infty}^{\infty}e^{zx}p(x)\,dx=e^{z^2/2}$ es una función analítica de z que se prolonga unívocamente por todo el plano complejo de la variable z. Colocando el valor it en lugar de z obtenemos, que la función característica buscada es

$$\varphi(t)=\int_{-\infty}^{\infty}e^{itx}p(x)\,dx=e^{-t^2/2},\ -\infty<t<\infty.$$

Señalemos que en la transformación lineal $\xi = \sigma\xi_0 + a$ la función característica $\varphi(t)$ de la magnitud ξ se obtiene de la función característica $\varphi_0(t)$ de la magnitud ξ_0 según la fórmula

$$\varphi(t) = \mathbf{M}e^{it(\sigma\xi+a)} = e^{ita}\varphi_0(\sigma t), \tag{5.10}$$

de modo que la función característica de la magnitud de Gauss ξ con el valor medio a y la dispersión σ^2 ($\xi = \sigma\xi_0 + a$, donde ξ_0 es la magnitud de Gauss con los parámetros (0,1)) es

$$\varphi(t) = e^{ita-\sigma^2t^2/2}, \quad -\infty < t < \infty. \tag{5.11}$$

Ejemplo (distribución simétrica exponencial y distribución de Cauchy). Examinemos la distribución de probabilidades con la densidad

$$p(x) = \frac{a}{2}e^{-a|x|}, \quad -\infty < x < \infty \tag{5.12}$$

(a es un determinado parámetro positivo). La función característica es

$$\varphi(t) = \frac{a}{2}\int_0^\infty (e^{itx} + e^{-itx})\,e^{ax}\,dx = \frac{a}{2}\int_0^\infty e^{(a+it)x}\,dx +$$

$$+ \frac{a}{2}\int_0^\infty e^{(a-it)x}\,dx = \frac{a}{2}\left(\frac{1}{a+it} - \frac{1}{a-it}\right) = \frac{a^2}{a^2+t^2}, \quad -\infty < t < \infty. \tag{5.13}$$

Indiquemos, que para $a = 1$ la función característica (5.13) se diferencia, de la densidad de probabilidad sólo por un factor positivo,

$$p(x) = \frac{1}{\pi}(1+x^2)^{-1}, \quad -\infty < x < \infty, \tag{5.14}$$

que se obtiene de la fórmula general (4.28) para la distribución de Student cuando $m = 1$. Esta densidad nos da la distribución de probabilidades de la relación de dos magnitudes de Gauss independientes con parámetros (0,1) y también la distribución de probabilidades de la magnitud tg ξ, donde ξ es la magnitud aleatoria, distribuida uniformemente en el segmento $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ y otras; la distribución con tal densidad de probabilidad se llama *distribución de Cauchy*.

Comparando a (5.12)—(5.14) y utilizando la fórmula de inversión (5.5), obtenemos para la función característica $\varphi(t)$ de la distribución de Cauchy, la expresión

$$\varphi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{itx}(1+x^2)^{-1}\,dx = 2\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itu}(1+u^2)^{-1}\,du\right] =$$

$$= e^{-|t|}, \quad -\infty < t < \infty, \tag{5.15}$$

Indiquemos una propiedad importante de las funciones características: *la suma de las magnitudes aleatorias independientes* $\xi = \xi_1 + \ldots + \xi_n$ *tiene la función característica* de la forma

$$\varphi(t) = \varphi_1(t) \ldots \varphi_n(t), \tag{5.16}$$

igual al producto de las funciones características $\varphi_1(t), \ldots, \varphi_n(t)$ de los sumandos aislados $\xi_1, \ldots, \xi_n$, que es un corolario de la propiedad general de las esperanzas matemáticas

$$Me^{it\xi} = M(e^{it\xi_1} \ldots e^{it\xi_n}) = (Me^{it\xi_1} \ldots (Me^{it\xi_n})$$

(véase la fórmula (4.6), cap. I).

Ejemplo (*distribución de triángulo*). Hallemos la función característica de la distribución de probabilidades con densidad

$$p(x) = \begin{cases} 1 - |x| & \text{para } |x| \leqslant 1, \\ 0 & \text{para } |x| > 1, \end{cases}$$

que coincide con la densidad de probabilidad de la suma $\xi = \xi_1 + \xi_2$ de las magnitudes aleatorias independientes ξ_1 y ξ_2 distribuidas uniformemente en los segmentos $[-1, 0]$ y $[0, 1]$ y que tienen las funciones características

$$\varphi_1(t) = \frac{1 - e^{-it}}{it} \text{ y } \varphi_2(t) = \frac{e^{it} - 1}{it}.$$

De acuerdo a la fórmula (5.16) la función característica buscada es

$$\varphi(t) = \varphi_1(t)\varphi_2(t) = \frac{|1 - e^{it}|^2}{t^2} = \frac{2(1 - \cos t)}{t^2}, \quad -\infty < t < \infty. \tag{5.17}$$

Ejemplo (*distribución binomial*). Hallemos la función característica de la distribución binomial o sea la distribución de la suma $\xi = \xi_1 + \ldots + \xi_n$ de magnitudes independientes igualmente distribuidas de la forma

$$\xi_k = \begin{cases} 1 & \text{con probabilidad } p, \\ 0 & \text{con probabilidad } 1 - p. \end{cases}$$

La función característica de cada sumando aislado es igual a $(1 - p) + e^{it}p$, y por consiguiente, la función característica buscada es

$$\varphi(t) = [1 + p(e^{it} - 1)]^n, \quad -\infty < t < \infty. \tag{5.18}$$

Ejemplo (χ^2-*distribución*). Hallemos la función característica de la magnitud aleatoria

$$\chi^2 = \xi_1^2 + \ldots + \xi_n^2.$$

o sea, de la suma de los cuadrados de magnitudes de Gauss independientes con parámetros (0, 1) que tienen la densidad de probabilidad la forma

$$p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/1}, \quad -\infty < x < \infty.$$

La densidad de probabilidad de cada una de las magnitudes ξ_k^2 tiene la forma (véase el ejemplo de la pág. 39)

$$p(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad 0 \leqslant y < \infty$$

($p(y) = 0$, para $y < 0$) y la función característica de cada sumando aislado ξ_k será

$$\int_0^\infty e^{ity} p(y)\, dy = \frac{1}{\sqrt{2\pi}} \int_0^\infty y^{-1/2} e^{-(1-2it)y/2}\, dy =$$

$$= (1-2it)^{-1/2} \frac{1}{\sqrt{\pi}} \int_0^\infty x^{-1/2} e^{-x}\, dx =$$

$$= (1-2it)^{-1/2} \frac{\Gamma(1/2)}{\sqrt{\pi}} = (1-2it)^{-1/2}, \qquad (5.19)$$

puesto que el valor de la función Γ en el punto 1/2 es igual a $\sqrt{\pi}$. La función característica de la χ^2-distribución con n grados de libertad será, por consiguiente,

$$\varphi(t) = (1 - 2it)^{-n/2}, \quad -\infty < t < \infty. \qquad (5.20)$$

Para $n > 1$ la función $|\varphi(t)|$, $-\infty < t < \infty$, es integrable y la densidad de la χ^2-distribución se podría hallar por la fórmula de inversión (5.5). Pero es más sencillo volver a las expresiones de (5.19). Se observa fácilmente, que si en lugar de $y^{-1/2}$ en la expresión inicial de la densidad $p(y)$, hubiese $y^{n/2-1}$, entonces, para el correspondiente factor normalizado D tendríamos

$$D \int_0^\infty e^{ity} y^{n/2-1} e^{-y/2}\, dy = (1-2it)^{-n/2}.$$

Ya que la función característica $\varphi(t) = (1 - 2it)^{-n/2}$ determina unívocamente la χ^2-distribución correspondiente, entonces, se debe concluir que esta distribución tiene la densidad de probabilidad

$$p_n(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, \quad 0 \leqslant x < \infty, \qquad (5.21)$$

donde el factor normalizador $D = \dfrac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}$ se determina de la condición

$$\int_0^\infty p_n(x)\,dx = D\int_0^\infty x^{n/2-1}e^{-x/2}\,dx = D\cdot 2^{n/2}\int_0^\infty y^{n/2-1}e^{-y}\,dy =$$

$$= D2^{n/2}\Gamma\left(\frac{n}{2}\right) = 1.$$

Ejemplo (estabilidad de la distribución de Cauchy). Sea $\xi_1, \xi_2, \ldots$, una sucesión de magnitudes aleatorias independientes que tienen la distribución de probabilidades de Cauchy. La función característica de cada una de ellas es (véase (5.15))

$$\varphi(t) = e^{-|t|}, \quad -\infty < t < \infty.$$

Examinemos la media aritmética

$$\xi = \frac{1}{n}\sum_{k=1}^{n}\xi_k.$$

Se ve fácilmente que la función característica $\varphi_\xi(t)$ de la magnitud ξ es

$$\varphi_\xi(t) = \varphi\left(\frac{t}{n}\right)^n = e^{-|t|},$$

y de tal modo, para cualquier n, la media aritmética $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ tiene la misma distribución de probabilidades, que cada una de las magnitudes ξ_k, $k = 1, 2, \ldots$ (Señalemos que las magnitudes indicadas no tienen esperanza matemática; compárese con la ley de los grandes números.)

2. Convergencia de las distribuciones de probabilidades. Se dice que la distribución de probabilidades, de las magnitudes aleatorias ξ_n, $n = 1, 2, \ldots$ *converge débilmente* hacia la distribución de probabilidades con la función de distribución $F(x)$, si

$$\mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} \to F(x'') - F(x') \quad \text{para} \quad n \to \infty \qquad (5.22)$$

para cualesquiera puntos x', x'' en los cuales $F(x)$ es continua.

Puede servir de ejemplo la convergencia débil de la distribución de sumas normalizadas $S_n^* = \dfrac{S_n - \mathbf{M}S_n}{\sqrt{\mathbf{D}S_n}}$, $n = 1, 2, \ldots$, en el esquema de pruebas de Bernoulli, hacia la ley normal con función de distribución $F(x) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} e^{-u^2/2}\,du$ (véase (3.6)).

Evidentemente, para cualquier punto de continuidad de $F(x)$ y para $x' = x'' = x$ de la relación (5.22) se deduce que

$$\mathbf{P}\{\xi_n = x\} \to 0 \quad \text{para} \quad n \to \infty,$$

y, de tal modo, la convergencia débil también significa que

$$\mathbf{P}\{x' < \xi_n \leqslant x''\} = F_{\xi_n}(x'') - F_{\xi_n}(x') \to F(x'') - F(x'). \quad (5.23)$$

(pra cualesquiera puntos de continuidad de $F(x)$).

Consideremos especialmente las magnitudes de números enteros, cuyos valores posibles son los enteros $k = 0, \pm 1, \ldots$ En este caso, evidentemente, la distribución límite de probabilidades deberá estar concentrada en los puntos de números enteros (es decir, $F(x) = \sum_{k \leqslant x} P(k)$, $\sum_k P(k) = 1$) y la relación (5.22) es equivalente a que

$$\mathbf{P}\{\xi_n = k\} \to P(k) \quad (5.24)$$

para cada k aislado ($P(k)$, $k = 0, \pm 1, \ldots$, es la distribución límite de probabilidades).

En efecto, en cualquier intervalo finito $[x', x'']$ cae, solamente, un número finito de puntos enteros, y de la relación (5.22) se deduce que

$$\mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} = \sum_{x' \leqslant k \leqslant x''} \{\mathbf{P}\xi_n = k\} \to \sum_{x' \leqslant k \leqslant x''} P(k) = F(x'') - F(x'),$$

por otra parte, $k - \varepsilon$, $k + \varepsilon$ para cualquier entero k y $0 < \varepsilon < 1$ son puntos de continuidad de la función $F(x) = \sum_{x \leqslant k} P(k)$, siendo $\mathbf{P}\{k - \varepsilon \leqslant \xi_n \leqslant k + \varepsilon\} = \mathbf{P}\{\xi_n = k\}$ y, por consiguiente, para una convergencia débil se deberá cumplir la relación límite (5.24).

Puede servir de ejemplo la convergencia de las distribuciones de magnitudes ξ_n o sea, el número de éxitos en las n pruebas de Bernoulli, hacia la distribución de probabilidades de Poisson, cuando para $n \to \infty$ el número medio de éxitos $a = np$ queda constante:

$$\mathbf{P}\{\xi_n = k\} \to \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, \ldots \text{ (véase (3.4)).}$$

Establezcamos algunos teoremas sobre la convergencia débil de la distribución de probabilidades.

Para mayor claridad introduzcamos la magnitud aleatoria ξ, cuya distribución de probabilidades figurará como distribución límite. Naturalmente, no se trata de ninguna convergencia de las propias magnitudes aleatorias ξ_n, $n = 1, 2, \ldots$, sino que se examinará sólo la convergencia de su distribución de probabilidades.

Teorema 1. *Si la sucesión de las distribuciones de probabilidades de las magnitudes* ξ_n, $n = 1, 2, \ldots$, *converge débilmente hacia la distribución de la magnitud* ξ, *entonces para cualquier función continua*

limitada $u = u(x)$, $-\infty < x < \infty$,

$$\lim_{n\to\infty} \mathbf{M}u(\xi_n) = \mathbf{M}u(\xi). \tag{5.25}$$

Demostración. En condiciones de la convergencia débil, las distribuciones examinadas de probabilidades son *compactas* en el *sentido, de que para cualquier* $\varepsilon > 0$ *se encuentra un tal valor finito* de a que

$$\mathbf{P}\{|\xi_n| \leqslant a\} > 1 - \varepsilon \tag{5.26}$$

a la vez para todos los $n = 1, 2, \ldots$

En efecto, tomando un intervalo lo suficientemente grande $[-a_0, a_0]$ para el cual $\mathbf{P}\{|\xi| \leqslant a_0\} \geqslant 1 - \frac{\varepsilon}{2}$ y $\mathbf{P}\{|\xi_n| \leqslant a_0\} \to$ $\to \mathbf{P}\{|\xi| \leqslant a_0\}$, obtenemos que para todos los valores suficientemente grandes de n ($n \geqslant n_0$) y $a \geqslant a_0$ se cumplirá la condición (5.26); y para cada número finito de los valores restantes $n = 1, \ldots, n_0$, con la correspondiente elección de a, también se cumplirá la condición (5.26).

Tomemos una función continua limitada arbitraria $u = u(x)$, $|u(x)| \leqslant K$. En el intervalo finito $[-a, a]$ se la puede aproximar con la exactitud que se quiera, con las funciones «escalonadas» del tipo

$$u_\varepsilon(x) = \begin{cases} u(x_k) & \text{para } x_{k-1} < x \leqslant x_k \ (k = 1, \ldots, N), \\ 0 & \text{para } x < -a, \ x > a, \end{cases}$$

donde $a = x_0 \leqslant x_1 \leqslant \ldots \leqslant x_N = a$ son los puntos de continuidad de la función $F(x)$ tales que $|u(x) - u(x_k)| \leqslant \varepsilon$ para $x_{k-1} < x \leqslant$ $\leqslant x_k$. De la condición de compacidad (5.26) se deduce que

$$|\mathbf{M}u(\xi_n) - \mathbf{M}u_\varepsilon(\xi_n)| \leqslant \mathbf{M}|u(\xi_n) - u_\varepsilon(\xi_n)| \leqslant$$
$$\leqslant K\mathbf{P}\{|\xi_n| > a\} + \varepsilon \leqslant (K + 1)\varepsilon$$

para todos los $n = 1, 2, \ldots$, y análogamente,

$$|\mathbf{M}u(\xi) - \mathbf{M}u_\varepsilon(\xi)| \leqslant (K + 1)\varepsilon.$$

Para cualquier valor fijado N, de la relación (5.23) obtenemos que

$$\mathbf{M}u_\varepsilon(\xi_n) = \sum_{k=1}^{N} u(x_k)\mathbf{P}\{x_{k-1} < \xi_n \leqslant x_k\} \to$$
$$\to \sum_{k=1}^{N} u(x_k)\mathbf{P}\{x_{k-1} < \xi \leqslant x_k\} = \mathbf{M}u_\varepsilon(\xi),$$

en conclusión obtenemos, que cualquiera que sea el valor tomado de antemano $\varepsilon > 0$,

$$\lim_{n\to\infty} |\mathbf{M}u(\xi_n) - \mathbf{M}u(\xi)| \leqslant 2(K + 1)\varepsilon,$$

de donde se deduce la relación (5.25).

Teorema 2. *Supongamos que la sucesión de magnitudes aleatorias* ξ_n, $n = 1, 2, \ldots$, converge en su media cuadrática hacia la mag-

nitud aleatoria ξ; entonces la distribución de probabilidades de estas magnitudes converge débilmente hacia la distribución límite de la magnitud ξ.

Demostración. Para cualesquiera ε y δ tan pequeños como se quiera y para los valores de n suficientemente grandes

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \leqslant \frac{\mathbf{M}|\xi_n - \xi|^2}{\varepsilon^2} \leqslant \delta.$$

En estas condiciones, evidentemente,

$$\mathbf{P}\{x' + \varepsilon \leqslant \xi \leqslant x'' - \varepsilon\} - \delta \leqslant \mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} \leqslant \leqslant \mathbf{P}\{x' - \varepsilon \leqslant \xi \leqslant x'' + \varepsilon\} + \delta,$$

de donde, para los puntos de continuidad x', x'' de la función de distribución $F_\xi(x)$ cuando $\varepsilon \to 0$, obtenemos

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} - \delta \leqslant \mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} \leqslant \leqslant \mathbf{P}\{x' \leqslant \xi \leqslant x''\} + \delta;$$

está claro que $\mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} \to \mathbf{P}\{x' \leqslant \xi \leqslant x''\}$ para $n \to \infty$.

Utilizando el teorema 2 deducimos la fórmula de inversión (5.6).

Examinemos, precisamente, la sucesión de magnitudes aleatorias $\xi_n = \xi + \Delta_n$, donde Δ_n son magnitudes aleatorias independientes de ξ que convergen en su media cuadrática hacia cero y que tienen las funciones características absolutamente integrables $\delta_n(t)$, $\int_{-\infty}^{\infty} |\delta_n(t)|\,dt < \infty$ (por ejemplo, éstas pueden ser magnitudes de Gauss con los parámetros $(0, \sigma_n)$, donde $\sigma_n \to 0$; en este caso $\delta_n(t) = e^{-\sigma_n^2 t^2/2}$, véase (5.11)). Las funciones características $\varphi_n(t) = \varphi(t)\cdot\delta_n(t)$ de las magnitudes $\xi_n = \xi + \Delta_n$, donde $\varphi(t)$ es la función característica de la magnitud que nos interesa ξ, son también integrables absolutamente, porque $|\varphi(t)| \leqslant 1$ y $|\varphi_n(t)| \leqslant |\delta_n(t)|$. Por consiguiente, las magnitudes ξ_n tienen densidades de probabilidad $p_n(x)$ que están ligadas con las funciones características $\varphi_n(t)$ a través de la transformación de Fourier (véase (5.5)):

$$p_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\varphi_n(t)\,dt.$$

Integrando por x, $x' \leqslant x \leqslant x''$ obtenemos que

$$\mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} = \int_{x'}^{x''} p_n(x)\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{x'}^{x''} e^{-itx}\,dx\right]\varphi_n(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ix''t} - e^{-ix't}}{it}\varphi(t)\,\delta_n(t)\,dt,$$

y según el teorema 2

$$\mathbf{P}\{x' \leqslant \xi \leqslant x''\} = \lim_{n\to\infty} \mathbf{P}\{x' \leqslant \xi_n \leqslant x''\} =$$

$$= \lim_{n\to\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ix''t} - e^{-ix't}}{-it} \varphi(t)\, \delta_n(t)\, dt \quad (5.27)$$

para cualesquiera puntos de continuidad de la función de distribución $F(x)$ de la magnitud ξ.

La relación límite (5.27) representa en sí una fórmula de inversión, cuyo caso particular es la fórmula (5.6).

Consideremos la conocida *igualdad de Parceval* para la transformación de Fourier:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_1(t)\, \varphi_2(t)\, dt = \int_{-\infty}^{\infty} p_1(x-y)\, p_2(y)\, dy, \quad (5.28)$$

donde $p_1(x)$, $p_2(x)$ son ciertas funciones absolutamente integrables, $\varphi_1(t)$, $\varphi_2(t)$ son sus transformaciones de Fourier (determinadas por la fórmula general (5.4)), con ello se supone que el producto $\varphi(t) = \varphi_1(t)\, \varphi_2(t)$, también es una función absolutamente integrable.

Detengámonos brevemente en la deducción de la igualdad (5.28). Como se puede comprobar fácilmente, $\varphi_1(t)\, \varphi_2(t)$ es la convolución de la transformación de Fourier

$$p(x) = \int_{-\infty}^{\infty} p_1(x-y)\, p_2(y)\, dy:$$

$$\int_{-\infty}^{\infty} e^{itx} \Big[\int_{-\infty}^{\infty} p_1(x-y)\, p_2(y)\, dy \Big] dx = \int_{-\infty}^{\infty} e^{itv} \Big[\int_{-\infty}^{\infty} e^{it(x-v)} \times$$

$$\times\, p_1(x-y)\, dx \Big] p_2(y)\, dy = \varphi_1(t) \int_{-\infty}^{\infty} e^{itv} p_2(y)\, dy = \varphi_1(t)\, \varphi_2(t),$$

de modo que en condiciones de absoluta integrabilidad de la función $\varphi(t) = \varphi_1(t)\, \varphi_2(t)$, la igualdad (5.28) no es otra cosa que la fórmula (5.5) para la transformación inversa de Fourier.

Para la función real $p_2(x)$ su transformación de Fourier $\varphi_2(t)$ será tal que $\varphi_2(-t) = \overline{\varphi_2(t)}$ y la igualdad de Parceval se podrá obtener en la siguiente forma:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \varphi_1(t)\, \overline{\varphi_2(t)}\, dt = \int_{-\infty}^{\infty} p_1(x+y)\, p_2(y)\, dy,$$

de donde para $x = 0$ y $p_1 = p_2 = p$, $\varphi_1 = \varphi_2 = \varphi$ obtenemos la llamada *identidad de Plancherel*:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi|^2 \, dt = \int_{-\infty}^{\infty} |p(x)|^2 \, dx \tag{5.29}$$

(se supone que

$$\int_{-\infty}^{\infty} |\varphi(t)|^2 \, dt < \infty).$$

De acuerdo con la identidad de Plancherel la distancia media cuadrática

$$\| p_1(x) - p_2(x) \| = \sqrt{\int_{-\infty}^{\infty} |p_1(x) - p_2(x)|^2 \, dx}$$

entre las funciones $p_1(x)$ y $p_2(x)$ se diferencia, sólo en un factor constante de la distancia del mismo tipo para sus transformaciones de Fourier $\varphi_1(t)$ y $\varphi_2(t)$; suponiendo, precisamente, en la fórmula (5.29) $p(x) = p_1(x) - p_2(x)$ y $\varphi(t) = \varphi_1(t) - \varphi_2(t)$ obtenemos

$$\int_{-\infty}^{\infty} |p_1(x) - p_2(x)|^2 \, dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi_1(t) - \varphi_2(t)|^2 \, dt. \tag{5.30}$$

Utilizando este hecho demostraremos la siguiente proposición.

Teorema 3. *La sucesión de distribuciones de probabilidades con las funciones características* $\varphi_n(t)$, $n = 1, 2, \ldots$, *converge débilmente hacia la distribución con la función característica* $\varphi(t)$, *entonces y sólo entonces, cuando*

$$\varphi_n(t) \to \varphi(t) \quad \text{para} \quad n \to \infty \tag{5.31}$$

uniformemente según t en cada intervalo finito.

Demostración. Supongamos que se trata de las distribuciones de probabilidades de las magnitudes aleatorias ξ_n, $n = 1, 2, \ldots$, que convergen débilmente hacia la distribución de la magnitud aleatoria ξ. Refiriéndonos a la demostración del teorema 1, observamos fácilmente, que para las funciones continuas $u(x) = e^{itx}$ (donde el parámetro t cambia en un determinado intervalo finito $|t| \leqslant T$)

$$\varphi_n(t) = \mathbf{M} e^{it\xi_n} \to \mathbf{M} e^{it\xi} = \varphi(t)$$

uniformemente según t, ya que

$$|e^{itx} - e^{itx_k}| \leqslant C |x - x_k|, \qquad k = 1, \ldots, N,$$

donde la constante C depende sólo de T. Por otra parte, si se ha cumplido la condición (5.31), entonces utilizando el mismo método que fue aplicado a la deducción de la fórmula de inversión (5.27), para las magnitudes aleatorias $\eta_n = \xi_n + \Delta$ y $\eta = \xi + \Delta$ con las funciones características $\psi_n(t) = \varphi_n(t)\,\delta(t)$ y $\psi(t) = \varphi(t)\,\delta(t)$ (donde $\delta(t) = e^{-\sigma^2 t^2/2}$ es la función característica de la magnitud de Gauss Δ, independiente de ξ_n y ξ) tendremos

$$\int_{-\infty}^{\infty} |\psi_n(t) - \psi(t)|^2\,dt = \int_{-\infty}^{\infty} |\varphi_n(t) - \varphi(t)|^2\,\delta(t)^2\,dt \to 0 \quad \text{para} \quad n \to \infty,$$

ya que para cualquier $\varepsilon > 0$ se encuentra un tal T que

$$\int_{|t|>T} |\varphi_n(t) - \varphi(t)|^2\,\delta(t)^2\,dt \leqslant 2 \int_{|t|>T} \delta(t)^2\,dt \leqslant \varepsilon,$$

y para $\varphi_n(t) \to \varphi(t)$ uniformemente según t, $|t| \leqslant T$,

$$\int_{-T}^{T} |\varphi_n(t) - \varphi(t)|^2\,\delta(t)^2\,dt \to 0.$$

Por consiguiente, para las densidades $p_n(x)$ y $p(x)$ de distribuciones de probabilidades de las magnitudes η_n y η

$$\int_{-\infty}^{\infty} |p_n(x) - p(x)|^2\,dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\psi_n(t) - \psi(t)|^2\,dt \to 0,$$

de donde se deduce la convergencia débil de estas distribuciones: para cualquier intervalo finito $[x', x'']$

$$\mathbf{P}\{x' \leqslant \eta_n \leqslant x''\} = \int_{x'}^{x''} p_n(x)\,dx \to \int_{x'}^{x''} p(x)\,dx = \mathbf{P}\{x' \leqslant \eta \leqslant x''\}.$$

Se ve fácilmente, que también tiene lugar la convergencia débil para las distribuciones de probabilidades de las magnitudes ξ_n, $n = 1, 2, \ldots$ (véase el teorema 2), ya que $\eta_n = \xi_n + \Delta$ y $\eta = \xi + \Delta$ se diferencian en su media cuadrática, de las magnitudes iniciales ξ_n y ξ todo lo pequeño que se desee $\sigma = \sqrt{\mathbf{M}\Delta^2}$. Esto es lo que se exigía demostrar.

Demostremos en conclusión *el teorema límite central* en la siguiente forma. Sea

$$S_n^* = \sum_{k=1}^{n} \xi_{kn}, \quad n = 1, 2, \ldots,$$

una sucesión de sumas de magnitudes independientes ξ_{kn}, $k = 1, \ldots, n$ tales que $M\xi_{kn} = 0$, $\sigma_{kn}^2 = M\xi_{kn}^2 \to 0$ para $n \to \infty$ uniformemente por k) $\sum_{k=1}^{n} \sigma_{kn}^2 = 1$. Entonces, *para la condición de Liapunov*

$$\sum_{k=1}^{n} M|\xi_{kn}|^3 \to 0 \text{ para } n \to \infty, \tag{5.32}$$

las distribuciones de probabilidades de las magnitudes S_n^*, $n = 1, 2, \ldots$, *convergen débilmente hacia la distribución normal standard*, a saber:

$$\lim_{n\to\infty} P\{x' \leqslant S_n^* \leqslant x''\} = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2}\, dx$$

(compárese con (4.17), donde figuran las magnitudes $\xi_{kn} = (\xi_k - a_k)/B_n$, $k = 1, 2, \ldots$).

Demostremos esto apoyándonos en el teorema 3. Demostremos precisamente, que las funciones características $\varphi_n(t)$ de las magnitudes S_n^* convergen hacia la función característica $\varphi(t) = e^{-t^2/2}$ de la distribución normal standard.

Para las funciones características $\varphi_{kn}(t)$ de los sumandos independientes ξ_{kn}, tenemos según la fórmula (5.8)

$$\varphi_{kn}(t) = 1 - \frac{\sigma_{kn}^2}{2} t^2 + \frac{R_{kn}}{6} t^3,$$

donde $|R_{kn}| \leqslant M|\xi_{kn}|^3$. Ya que $\sigma_{kn}^2 \to 0$, $R_{kn} \to 0$ para $n \to \infty$ uniformemente por k, entonces $\varphi_{kn}(t) \to 1$ uniformemente por k y t en cada intervalo finito; en particular, $|\varphi_{kn}(t) - 1| \leqslant \frac{1}{2}$ para valores de n suficientemente grandes y tiene sentido hablar sobre el logaritmo de las funciones, $\varphi_{kn}(t)$. Para $\varphi_n(t) = \prod_{k=1}^{n} \varphi_{kn}(t)$ cuando $n \to \infty$, tenemos

$$\ln \varphi_n(t) = \sum_{k=1}^{n} \ln \varphi_{kn}(t) = \sum_{k=1}^{n} \ln\left(1 - \frac{\sigma_{kn}^2}{2} t^2 + \frac{R_{kn}}{6} t^3\right) \sim$$

$$\sim \sum_{k=1}^{n} \left(-\frac{\sigma_{kn}^2}{2} t^2 + \frac{R_{kn}}{6} t^3\right) = -\frac{t^2}{2} + \left(\sum_{k=1}^{n} R_{kn}\right) \frac{t^3}{6} \sim -\frac{t^2}{2}$$

uniformemente por t en cada intervalo finito, ya que para la condición (5.32)

$$\left|\sum_{k=1}^{n} R_{kn}\right| \leqslant \sum_{k=1}^{n} M|\xi_{kn}|^3 \to 0.$$

Por consiguiente,

$$\varphi_n(t) \to e^{-t^2/2} \text{ para } n \to \infty$$

uniformemente en cada intervalo finito, que es lo que se exigía demostrar.

CAPITULO III

ALGUNOS MODELOS DE PROCESOS ALEATORIOS

§ 1. ALGUNAS DEFINICIONES Y EJEMPLOS

1. Definición general del proceso aleatorio. Hablando sobre el proceso aleatorio, como regla, se toma en consideración una determinada magnitud aleatoria $\xi(t)$, que varía en el transcurso del tiempo t. Llamaremos *proceso aleatorio* $\xi = \xi(t)$ a la función de parámetro real $t \in T$, cuyos valores $\xi(t)$ para cada t son magnitudes aleatorias.

Las leyes del proceso aleatorio $\xi(t)$, $t \in T$ se determinan por las distribuciones conjuntas de probabilidades de sus valores $\xi(t_1)$, ... , $\xi(t_n)$ para los distintos t_1, . . ., t_n (ellas se llaman *distribuciones de dimensión finita* del proceso aleatorio dado).

Por ejemplo, éstas pueden ser distribuciones de Gauss con la densidad de probabilidad (véase (4.10), cap. II)

$$p_{t_1, \ldots, t_n}(x_1, \ldots, x_n) =$$

$$= \frac{1}{(2\pi)^{n/2} |B|^{1/2}} \exp \left\{ -\frac{1}{2} \sum_{k, j=}^{n} b_{kj} [x_k - A(t_k)] [(x_j - A(t_j)] \right\},$$

donde $A(t_1)$, ... $A(t_n)$ es el valor de la función

$$A(t) = \mathbf{M}\xi(t), \quad t \in T,$$

llamada *valor medio* del proceso aleatorio $\xi(t)$, $t \in T$ y la matriz de correlación $B = \{B(t_k, t_j)\}$ está compuesta de los valores de la función

$$B(s, t) \mathbf{M}[\xi(s) - A(s)][\xi(t) - A(t)], \quad s, t \in T$$

llamada *función de correlación* de este proceso (recordemos que los coeficientes $b_{kj} k, j = 1, \ldots, n$ bajo el signo del exponente forman la matriz $\{b_{kj}\}$, inversa a la matriz de correlación $\{B(t_k, t_j)\}$ de las magnitudes aleatorias $\xi(t_1)$, . . ., $\xi(t_n)$; el propio proceso aleatorio $\xi(t)$, $t \in T$ con distribuciones de Gauss de dimensiones finitas, también se llama *distribución de Gauss*.

Cada valor $\xi(t)$ del proceso aleatorio, siendo una magnitud aleatoria, depende formalmente del resultado elemental ω: $\xi(t) = \xi(\omega, t)$. Examinando el proceso aleatorio $\xi(t)$, $t \in T$, para cada resultado aleatorio aislado ω, nosotros tratamos la función correspondiente $\xi(\omega, \cdot) = \xi(\omega, t)$ del parámetro $t \in T$, que se denomina *trayectoria* o *función seleccionada* del proceso aleatorio dado. Observando realmente el proceso aleatorio, observamos de hecho una de sus posibles trayectorias $x = x(t)$, $t \in T$.

Podemos imaginarnos que se tiene un conjunto determinado X de todas las trayectorias posibles $x = x(t)$, $t \in T$, y un «mecanismo aleatorio» elige una de estas funciones $x \in X$. Con ello la elección aleatoria de tal o cual trayectoria $x = x(t)$, $t \in T$, de X se puede considerar como un resultado elemental. Recordemos que tal punto de vista ya fue expresado antes por nosotros en relación a las magnitudes aleatorias (la magnitud aleatoria n-dimensional se puede considerar como una función aleatoria $\xi(t)$ del parámetro t, que recorre un número finito de valores $t = 1, \ldots, n$). Se indicó precisamente, que examinando las magnitudes $[\xi(1), \ldots, \xi(n)]$, se puede considerar espacio de resultados elementales al espacio X de n dimensiones de vectores $x = [x(1), \ldots, x(n)]$, tomando como resultado elemental aquel valor de $x = [x(1), \ldots, x(n)]$ que toma, según el caso, la magnitud observada $\xi = [\xi(1), \ldots, \xi(n)]$. Con ello el mecanismo aleatorio correspondiente, se da de hecho, por la distribución conjunta de probabilidades de las magnitudes $[\xi(1), \ldots, \xi(n)]$.

2. Los procesos aleatorios de Markov. Se denomina proceso aleatorio *de Markov* $\xi = \xi(t)$ si las magnitudes aleatorias $\xi(t)$, $t \geqslant u$, no dependen de $\xi(s)$, $s \leqslant u$, en cualquier momento de tiempo u para el valor fijado $\xi(u) = x$ (cualquiera que fuese x). Si acordamos considerar $\xi(t)$ como el estado *físico* de un sistema *físico* determinado en el momento de tiempo t, entonces el proceso de Markov $\xi = \xi(t)$, que describe la evolución del sistema examinado, se puede caracterizar más demostrativamente de la siguiente forma: el comportamiento del sistema después de un determinado momento de tiempo u en el estado conocido $x = \xi(u)$ no depende de su comportamiento hasta este momento.

Una propiedad característica de los procesos aleatorios de Markov se ve claramente en un ejemplo tan sencillo como es el juego infantil de «quien va despacio va lejos». En este juego la ficha del jugador deberá pasar un número determinado de puntos 1, 2, . . . El paso de un punto al otro se determina cada vez por el resultado del lanzamiento del dado de juego. Precisamente, si en el paso dado n, la ficha se encuentra en el punto $\xi(n) = i$, por las reglas del juego se establece el punto $\xi(n + 1)$ de su desplazamiento en el paso siguiente, en dependencia del número de puntos con que cayó el dado de juego. Desde cualquier punto i la ficha pasa en su paso siguiente al punto j correspondiente con una determinada probabilidad p_{ij}. Se comprende

fácilmente que el proceso aleatorio del cambio de estado $\xi(n)$ durante el «tiempo» $n = 0, 1, 2, \ldots$ es un proceso de Markov. (Aquí figura, en lugar del tiempo real, el número n de pasos realizados: se puede considerar convencionalmente, que cada paso se realiza en la unidad de tiempo).

Ejemplo (fluctuación aleatoria). Examinemos la fluctuación aleatoria de la partícula por los puntos de números enteros de la recta real, durante la cual la partícula se desplaza en cada paso en una unidad a la derecha con la probabilidad p y en una unidad a la izquierda con la probabilidad $q = 1 - p$. Sea $\xi(n)$ la posición de la partícula después de n pasos. La sucesión $\xi(0), \xi(1), \xi(2), \ldots$ es de Markov: si la partícula se encuentra en un determinado punto i, entonces su futuro comportamiento no depende de las circunstancias, precedentes a la caída en el punto i, y en los n pasos siguientes la partícula pasa al estado correspondiente j con la probabilidad $p_{ij}(n)$.

Está claro, que el paso de i a j, cuando $|i - j| > n$, es imposible y en este caso $p_{ij}(n) = 0$. También está claro, que la partícula durante n pasos, sólo puede pasar a aquellos estados j, para los cuales la diferencia $|i - j|$ tiene la misma paridad que n, es decir, el número

$$m = \frac{n + |j - i|}{2}$$

deberá ser entero. Para $j \geqslant i$ se puede caer en el estado j entonces y solamente entonces cuando entre todos los n pasos se realizan justamente $m = \frac{n + (j - i)}{2}$ pasos en dirección positiva. La probabilidad de esto es (véase (3.1), cap. II)

$$p_{ij}(n) = C_n^m p^m q^{n-m}, \quad j \geqslant i. \tag{1.1}$$

Análogamente se expresa la probabilidad del paso de i a j, para $j \leqslant i$:

$$p_{ij}(n) = C_n^m p^{n-m} q^m, \quad j \leqslant i.$$

Ejemplo (desintegración radiactiva) (véanse las págs. 87, 90). Consideraremos como anteriormente, que cada átomo Ra independiente de las circunstancias precedentes se transforma en el tiempo t en átomo Rn con la probabilidad $p(t) = 1 - e^{-\lambda t}$ $\left(\lambda = \frac{\log 2}{T}, T \text{ es la constante de semidesintegración}\right)$. Entonces, el número total $\nu(t)$ de átomos de Ra desintegrados en el tiempo t, igual al número de partículas α emitidas en este período de tiempo, está distribuido según la ley de Poisson:

$$\mathbf{P}\{\nu(t) = k\} = \frac{a^k}{k!} e^{-a}, \quad k = 0, 1, 2, \ldots,$$

donde $a = np\ (t)$, n es el número inicial de átomos de Ra. El número de átomos restantes de Ra será $\xi\ (t) = n - \nu\ (t)$. Si es conocida la cantidad de radio en un determinado momento s: $\xi\ (s) = i$, entonces, independientemente del carácter del proceso de desintegración hasta el momento s, se emiten k partículas α en intervalo desde s hasta t con la probabilidad $\frac{a^k}{k!} e^{-a}$, $a = ip\ (t - s)$. De tal modo, $\xi\ (t)$ es un proceso aleatorio de Markov, con ello la probabilidad de paso del estado $\xi\ (s) = i$ al estado $\xi\ (t) = j\ (t > s)$ es

$$p_{ij}(t-s) = \frac{a^{i-j}}{(i-j)!} e^{-a}, \quad a = ip\,(t-s), \quad j \leqslant i \qquad (1.2)$$

(para $j > i$, es imposible el paso de i a j).

Ejemplo (*proceso de Poisson*). Examinemos la corriente de sucesos de Poisson, en la cual el número de sucesos durante un determinado intervalo de tiempo (s, t) no depende del número de sucesos ocurridos hasta el momento s (véase (2.5) y más adelante, cap. II). Está claro, que el proceso aleatorio de la variación $\xi\ (t)$, del número de sucesos ocurridos hasta el momento actual t, es proceso de Markov, con ello la probabilidad del paso del estado $\xi\ (s) = i$ al estado $\xi\ t) = j(t > s)$es

$$p_{ij}\ (t-s) = \frac{(\lambda\,(t-s)^{j-i}}{(j-i)^{\cdot}} e^{-\lambda_{\mathfrak{s}}(t-s)} \quad j \leqslant i \qquad (1.3)$$

(para $j < i$, es imposible el paso de i a j).

Ejemplo (*movimiento browniano*). El proceso del movimiento browniano fue descrito en el p. 2 del § 3, cap. II). Como se ve de esta descripción (especialmente, si nos volvemos al modelo discreto), el movimiento de la partícula browniana desde cualquier estado $\xi\ (s) = x$ no depende de su comportamiento hasta el momento s (en el modelo discreto ella se traslada en el siguiente paso a uno de los estados $x + \Delta x$, $x - \Delta x$ con la misma probabilidad), con ello, la distribución de probabilidades de la magnitud $\xi\ (t)$, o sea, las coordenadas de la partícula browniana en el momento de tiempo $t > s$, para la condición de que $\xi\ (s) = x$, tiene la densidad

$$p\,(t-s,\ x,\ y) = \frac{1}{\sqrt{2\pi\,(t-s)}\,\sigma} e^{-(y-x)^2/2\,(t-s)\,\sigma^2}, \quad -\infty < y < \infty. \quad (1.4)$$

Examinemos el proceso aleatorio de Markov $\xi = \xi\ (t)$, considerando, como convenimos anteriormente, en que él describe el comportamiento de un determinado sistema físico, en el transcurso del tiempo t, y $\xi\ (t)$ significa el estado fásico de este sistema en el momento t.

Supongamos que se tiene un número finito o numerable de estados fásicos distintos. Si el sistema se encuentra en el estado $\xi\ (s) = x$,

en el momento de tiempo s, entonces después del tiempo $t - s$ se encontrará en tal o cual estado y con una probabilidad determinada, que designamos por $P(s, x, t, y)$; notemos, que para el proceso de Markov la probabilidad indicada no depende de su comportamiento hasta el momento s. Formalmente, $P(s, x, t, y)$ significa la probabilidad condicional de que $\xi(t) = y$ para la condición $\xi(s) = x$, $s \leqslant t$. Se puede decir que $P(s, x, t, y)$ es la probabilidad del paso desde el estado x, en que se encuentra el sistema en el momento s, al estado y durante el tiempo $t - s$.

En general, en el caso del espacio arbitrario de estados, se introducen las llamadas *probabilidades de paso* $P(s, x, t, B)$, o sea, las probabilidades del paso del estado inicial $x = \xi(s)$ a uno de los estados y del conjunto indicado de estados B durante el tiempo $t - s$ (formalmente, $P(s, x, t, B)$ significa la probabilidad condicional de que $\xi(t) \in B$ para la condición $\xi(s) = x$, $s \leqslant t$). Por ejemplo si sólo se tiene un conjunto finito o numerable de estados, entonces

$$P(s, x, t, B) = \sum_{y \in B} P(s, x, t, y);$$

si $\xi(t)$ es un vector aleatorio en el espacio de n dimensiones E^n, y su distribución de probabilidades en condiciones que $\xi(s) = x$ tiene la densidad $p(s, x, t, y)$, $y \in E^n$, entonces

$$P(s, x, t, B) = \int_B p(s, x, t, y)\, dy$$

(la densidad condicional de probabilidad $p(s, x, t, y)$, $y \in E^n$, la denominan corrientemente *densidad de paso* del proceso de Markov $\xi = \xi(t)$).

Las probabilidades de paso $P(s, x, t, B)$ del proceso de Markov en el espacio vectorial de n dimensiones E^n, dan de hecho las distribuciones de probabilidades del incremento $\xi(t) - \xi(s)$ para el valor fijado $\xi(s) = x$; si designamos, precisamente, por $B + x$ al conjunto de puntos $y \in E^n$, tales que $y - x \in B$, entonces para la condición $\xi(s) = x$, $\xi(t) - \xi(s) \in B$ con la probabilidad $P(s, x, t, B + x)$. Con respecto a esto señalemos una clase importante de *procesos aleatorios con incrementos independientes*, que poseen tales propiedades que los incrementos $\xi(t) - \xi(t_0)$ son independientes de las magnitudes $\xi(s)$, $s \leqslant t_0$ (cualesquiera que sean $s \leqslant t_0 \leqslant t$). Como ejemplo de procesos unidimensionales de tal tipo puede servir el proceso de Poisson o el movimiento browniano y conocidos por nosotros.

El proceso de Markov $\xi = \xi(t)$ se llama *homogéneo* si la regularidad de su comportamiento en cualquier intervalo (s, t) para el estado determinado $\xi(s) = x$, no depende de la situación del intervalo

(s, t) en el eje del tiempo; para las probabilidades de paso $P(s, x, t, B)$, esto significa de hecho que sólo depende de la diferencia $t - s$, y no de s y t, como en el caso general:

$$P(s, x, t, B) = P(t - s, x, B)$$

(todos los procesos de Markov expuestos anteriormente en los ejemplos, son homogéneos).

Naturalmente, la esencia de la definición no depende de las designaciones que, para comodidad de anotación, pueden ser cambiadas (compárense las probabilidades de paso en los ejemplos (1.1)—(1.3) o la densidad de paso en el ejemplo (1.4)).

En los ejemplos examinados anteriormente, los procesos de Markov, poseen la tal llamada p r o p i e d a d r i g u r o s a d e M a r k o v, que se puede describir en términos generales del siguiente modo: al caer en tal o cual estado x, el comportamiento ulterior del proceso (para el valor dado de x) no depende de su comportamiento en el pasado.

Precisemos este enunciado, utilizando el concepto del momento aleatorio de tiempo i n d e p e n d i e n t e d e l f u t u r o Así se llama cualquier magnitud τ acerca de la cual se puede unívocamente decir, para cualquier t de la trayectoria $\xi(s)$, $s \leqslant t$, si es $\tau \leqslant t$ o es $\tau > t$ (τ también se llama *momento de tiempo de Markov*). Como ejemplo de magnitudes independientes del futuro puede servir el momento τ de la primera caída en tal o cual estado x, y en general, el momento de caída en un estado determinado x, después de algún momento τ_0 de Markov.

Para aclaración, exponemos el ejemplo del momento aleatorio τ, d e p e n d i e n t e del futuro y que no es de Markov: $\tau = \tau_1$ si el momento τ_1 de la primera caída después de t_0, en un determinado estado x es menor que $t = t_0 + h$, y $\tau = \tau_1 - h$, si $\tau_1 > t$; está claro que en la trayectoria $\xi(s)$, $t_0 \leqslant s \leqslant t$, hablando en general, no se puede determinar si es $\tau \leqslant t$ (es decir, $\tau_1 \leqslant t + h$) o no. El proceso de Markov $\xi = \xi(t)$ se llama *proceso riguroso de Markov*, si para cualquier momento de tiempo de Markov τ, a condición de que $\xi(\tau) = x$, donde x es el estado conocido, el comportamiento del proceso para $t > \tau$ no depende de su comportamiento anterior al momento τ, con ello el paso desde $x = \xi(\tau)$ a tal o cual estado y durante el tiempo h se realiza con las mismas probabilidades que en el caso del momento fijado $\tau = s$; más exacto: la probabilidad condicional durante el tiempo h, resulta ser en el conjunto B igual a $P(\tau, x, \tau + h, B)$, donde $P(s, x, t, B)$ es la probabilidad de paso del proceso examinado de Markov.

Al hablar en lo sucesivo sobre el proceso de Markov $\xi = \xi(t)$ (véanse las cadenas de Markov, etc.) sobreentenderemos que se cumple la propiedad rigurosa de Markov descrita anteriormente.

§ 2. CADENAS DE MARKOV. CLASIFICACION DE LOS ESTADOS. DISTRIBUCIONES ESTACIONARIAS

1. Probabilidades de paso. Como convenimos anteriormente, al examinar el proceso aleatorio $\xi = \xi(t)$, hablaremos del sistema, cuyo estado físico en el momento de tiempo t es $\xi(t)$.

Sea $\xi(t)$, $t \geqslant 0$ un proceso homogéneo de Markov con un número finito o numerable de estados para los cuales se pueden tomar los números naturales $i = 1, 2, \ldots$

Supongamos que el parámetro t toma sólo valores de números enteros $t = 0, 1, 2, \ldots$; entonces, tratamos sobre la cadena de pasos

$$\xi(0) \to \xi(1) \to \xi(2) \to \ldots$$

El proceso $\xi = \xi(t)$ del tipo descrito se llama corrientemente *cadena de Markov*.

Supongamos que están dadas las probabilidades de paso p_{ij}, o sea, las probabilidades de paso del sistema del estado i al estado j en un paso, en otras palabras, la probabilidad de que el sistema se encuentre en tal o cual estado $j = 1, 2, \ldots$ en condiciones de que el sistema estuviese en el paso anterior en el estado i:

$$p_{ij} = \mathbf{P}\{\xi(n) = j \mid \xi(n-1) = i\} \left(\sum_j p_{ij} = 1; \quad i = 1, 2, \ldots\right).$$

Supongamos también que el sistema se encuentra en el momento inicial en uno de los estados i con la probabilidad correspondiente p_i^0, $\sum_i p_i^0 = 1$ (en particular, si $p_i^0 = 1$, $p_k^0 = 0$ para $k \neq i$, entonces, esto significa que i es el estado inicial). ¿Cuál es la probabilidad de que el sistema se encuentre después de n pasos, en tal o cual estado $j = 1, 2, \ldots$?

Designemos a esta probabilidad por $p_j(n)$: $p_j(n) = \mathbf{P}\{\xi(n) = j\}$. Después de $n - 1$ pasos el sistema se encontrará obligatoriamente en uno de los estados $k = 1, 2, \ldots$, con esto se encontrará en k con la probabilidad designada por $p_k(n-1)$. Para la condición de que después de $n - 1$ pasos se encuentre el sistema en el estado k, la probabilidad de encontrarse después de n pasos en el estado j será igual a la probabilidad p_{kj} de paso k a j. Utilizando la fórmula de la probabilidad total, obtenemos

$$\mathbf{P}\{\xi(n) = j\} = \sum_k \mathbf{P}\{\xi(n) = j \mid \xi(n-1) = k\} \mathbf{P}\{\xi(n-1) = k\}.$$

Esto da la siguiente relación recurrente para las probabilidades $p_j(n)$, $j = 1, 2, \ldots$:

$$p_j(0) = p_j^0, \quad p_j(n) = \sum_k p_k(n-1) \cdot p_{kj}, \quad (n = 1, 2, \ldots). \quad (2.1)$$

Si el sistema se encuentra en el momento inicial en un estado determinado i, entonces la distribución inicial de probabilidades es $p_i^0 = 1$, $p_k^0 = 0$ para $k \neq i$ y la probabilidad $p_j(n)$ coincide con la probabilidad $p_{ij}(n)$ de que el sistema pase del estado i, en n pasos, al estado j:

$$p_{ij}(n) = \mathbf{P}\{\xi(n) = j \mid \xi(0) = i\}, \quad i, j = 1, 2, \ldots$$

Para la distribución inicial de la forma $p_i^0 = 1$, $p_k^0 = 0$, para $k \neq i$, la fórmula (2.1) da las siguientes relaciones entre *las probabilidades de paso* $p_{ij}(n)$; $i, j = 1, 2, \ldots$:

$$p_{ij}(0) = \begin{cases} 1 \text{ para } j = i \\ \qquad\qquad p_{ij}(n) = \sum_k p_{ik}(n-1)\, p_{kj}, \\ 0 \text{ para } j \neq i \end{cases} \quad (2.2)$$

$$(n = 1, 2, \ldots).$$

Es cómodo introducir la matriz $P(n) = \{p_{ij}(n)\}$. De acuerdo a la fórmula (2.2)

$$P(0) = I, \quad P(1) = P, \quad P(2) = P(1)\, P = P^2, \ldots,$$

donde I es la matriz unitaria, $P = \{p_{ij}\}$ es la matriz de las probabilidades de paso. Vemos que tiene lugar la siguiente igualdad

$$P(n) = P^n \quad (n = 0, 1, 2, \ldots).$$

Conociendo las probabilidades de paso $p_{ij}(n)$ (véanse (2.1), (2.2)), se puede determinar fácilmente, *la distribución de dimensiones finitas* de la cadena de Markòv examinada. Precisamente, para cualesquiera momentos de tiempo $n_1 \leqslant n_2 \leqslant \ldots \leqslant n_k$ y de estados $i_1, i_2, \ldots, i_k$ en condiciones de que $\xi(n_1) = i_1 \ldots, \xi(n_{k-1}) = i_{k-1}$, la probabilidad de encontrarse en el estado i_k en el momento n_k será igual a $p_{i_{k-1} i_k}(n_k - n_{k-1})$, o de otro modo

$$\mathbf{P}\{\xi(n_1) = i_1, \ldots, \xi(n_k) = i_k \mid \xi(n_1) = i_1, \ldots, \xi(n_{k-1}) = i_{k-1})\} = p_{i_{k-1} i_k}(n_k - n_{k-1});$$

de donde

$$\mathbf{P}\{\xi(n_1) = i_1, \ldots, \xi(n_k) = i_k\} = \mathbf{P}\{\xi(n_1) = i_1, \ldots, \xi(n_{k-1}) = i_{k-1}\} \cdot p_{i_{k-1} i_k}(n_k - n_{k-1});$$

y, análogamente,

$$\mathbf{P}\{\xi(n_1) = i_1, \ldots, \xi(n_{k-1}) = i_{k-1}\} =$$
$$= \mathbf{P}\{\xi(n_1) = i_1, \ldots, \xi(n_{k-2}) = i_{k-2}\}\, p_{i_{k-2}i_{k-1}}(n_{k-1} - n_{k-2}),$$

. . . . . .

Como resultado obtenemos la fórmula siguiente:

$$\mathbf{P}\{\xi(n_1) = i_1, \xi(n_2) = i_2, \ldots, \xi(n_k) = i_k\} =$$
$$= p_{i_1}(n_1) \cdot p_{i_1 i_2}(n_2 - n_1) \cdot \ldots \cdot p_{i_{k-1} i_k}(n_k - n_{k-1}). \qquad (2.3)$$

Examinemos el proceso homogéneo de Markov $\xi = \xi(t)$, cuyos estados varían durante el tiempo c o n t i n u o t, $t \geqslant 0$. Designemos por ξ_n el n-ésimo estado, según el recuento, y por τ_n, el momento de caída en el estado ξ_n (de otro modo, τ_n es el tiempo hasta el n-ésimo paso $\xi_{n-1} \to \xi_n$, $n = 1, 2, \ldots$); supongamos que $\tau_0 = 0$.

Está claro que el momento τ_n no depende del futuro, y se ve fácilmente, que gracias a la propiedad rigurosa de Markov del proceso inicial $\xi(t)$, el proceso de los pasos de estado en estado, descrito por la cadena

$$\xi_0 \to \xi_1 \to \xi_2 \to \ldots,$$

es una cadena de Markov. Hallemos las probabilidades de paso π_{ij} de esta cadena de Markov $\xi_n = \xi(\tau_n)$, $n = 0, 1, \ldots$

Hallemos al principio las distribuciones de probabilidades de las magnitudes $\tau_{n+1} - \tau_n$ para cada uno de los estados posibles $\xi_n = 1, 2, \ldots$. Debido a la homogeneidad del proceso $\xi = \xi(t)$, para la condición $\xi_n = i$, la distribución $\tau_{n+1} - \tau_n$ es igual que la distribución $\tau = \tau_i$ para un mismo estado inicial $i = \xi_0$ (τ no es otra cosa que el tiempo hasta el cambio del estado i).

Señalemos que, de acuerdo con la propiedad rigurosa de Markov, el comportamiento del proceso $\xi = \xi(t)$ después del momento τ_n, para cualquier $i = \xi(\tau_n)$, no depende de su comportamiento hasta este momento, y en particular, el tiempo $\tau_{n+1} - \tau_n$ de estancia en los estados $i = \xi(\tau_n)$ no depende de las magnitudes $\tau_{k+1} - \tau_k$, $k \leqslant n - 1$ (de donde, a propósito, no es difícil deducir que las diferencias indicadas $\tau_{n+1} - \tau_n$, $n = 0, 1, \ldots$, son magnitudes aleatorias independientes).

Supongamos que, para algún valor de s, el proceso se encuentra en un estado determinado i ($\xi(s) = i$). Examinemos el tiempo τ (después del momento s) hasta el momento de cambio de este estado i. Si $\tau > t$ y, en particular, el estado del proceso en el momento de tiempo $s_1 = s + t$ es el mismo: $\xi(s_1) = i$, entonces el tiempo τ_1 (después de s_1) hasta el momento de cambiar el estado, tiene una distribución de probabilidades igual que τ, ya que gracias a la homogeneidad del proceso, su comportamiento después del momento s_1 se somete a las mismas leyes, que después del momento s, (en ambos casos el

estado inicial es i). Esto significa que

$$\mathbf{P}\{\tau > u + v \mid \xi(s) = i, \quad \tau > u\} = \\ = \mathbf{P}\{\tau_1 > v \mid \xi(s_1) = i\} = \mathbf{P}\{\tau > v \mid \xi(s) = i\},$$

por que el suceso $\{\xi(s) = i, \ \tau > u\}$ incluye en sí al suceso $\{\xi(s_1) = i\}$, y con la condición de que $\xi(s_1) = i$, el comportamiento del proceso de Markov $\xi(t)$ no depende de su comportamiento hasta el momento s_1. Aquí obtenemos, como al deducir la fórmula (2.8), cap. II, que *el tiempo τ de espera hasta el cambio de estado tiene la distribución exponencial de probabilidades*:

$$\mathbf{P}\{\tau > t \mid \xi(s) = i\} = e^{-\lambda_i t}, \quad t \geqslant 0, \tag{2.4}$$

donde λ_i es una constante determinada no negativa ($1/\lambda_i = \mathbf{M}\tau$ es el tiempo medio de espera); *la constante λ_i se llama densidad de salida* desde el estado correspondiente i.

Para $\lambda_i = 0$, el proceso $\xi(t)$ se queda para siempre en el estado i; tal estado se llama *absorbente*; para $0 < \lambda_i < \infty$ la probabilidad de que el estado i se cambie en el pequeño intervalo Δt es

$$\lambda_i \Delta t + o(\Delta t), \tag{2.5}$$

donde $o(\Delta t)$ es un infinitamente pequeño de orden superior en relación a Δt, $\Delta t \to 0$.

Volviendo a la cuestión sobre las probabilidades de paso π_{ij}, supongamos que el paso del proceso $\xi(t)$ del estado i al estado j ($j \neq i$) en el tiempo pequeño Δt es

$$\lambda_{ij} \Delta t + o(\Delta t),$$

donde $o(\Delta t)/\Delta t \to 0$ para $\Delta t \to 0$.

Si la densidad de salida λ_i es igual a cero (el estado i es absorbente), el sistema después de caer en i, se queda allí para siempre, es decir, $\pi_{ij} = 0$ para todos los $j \neq i$; y si $\lambda_i \neq 0$, el sistema sale del estado i con la probabilidad 1 después de un tiempo finito τ (recordemos, que $\mathbf{P}\{\tau < t\} = 1 - e^{-\lambda_i t}$). Para la condición $\tau > t$ tenemos $\xi(t) = i$ y la probabilidad de paso en el intervalo $(t, \ t + \Delta t)$ será, según la suposición, $\lambda_{ij}\Delta t + o(\Delta t)$; por consiguiente, la probabilidad incondicional $\pi_{ij}(t, \ t + \Delta t)$ de tal paso es

$$\pi_{ij}(t, \ t + \Delta t) = [\lambda_{ij}\Delta t + o(\Delta t)] \cdot \mathbf{P}\{\tau > t\} = \\ = [\lambda_{ij}\Delta t + o(\Delta t)]\, e^{-\lambda_i t}.$$

Dividiendo el semieje $0 \leqslant t < \infty$ con los puntos $t = k\Delta t$, $k = 0, 1, \ldots$, tenemos

$$\pi_{ij} = \sum_{k=0}^{\infty} \pi_{ij}(k\Delta t, \ (k+1)\Delta t) = \\ = \sum_{k=0}^{\infty} [\lambda_{ij}\Delta t + o(\Delta t)]\, e^{-\lambda_i k \Delta t} \sim \lambda_{ij} \int_0^{\infty} e^{-\lambda_i t}\, dt,$$

de donde obtenemos

$$\pi_{ij} = \lambda_{ij} \int_0^\infty e^{-\lambda_i t}\, dt = \frac{\lambda_{ij}}{\lambda_i}, \quad j \neq i. \tag{2.6}$$

(Recordemos que examinamos el proceso $\xi_n = \xi(\tau_n)$, $n = 0, 1, \ldots$, de los pasos sucesivos, en el cual cada estado siguiente j es diferente del estado anterior i, de modo que $\pi_{ii} = 0$).

2. Estados reversibles e irreversibles. Examinemos la cadena de Markov con los estados 1, 2, ... y con las probabilidades $p_{ij}(n)$ de paso en n pasos, del estado i al estado j; $i, j = 1, 2, \ldots$

Supongamos que en el momento inicial el sistema se encuentra en un estado determinado i; Designemos por v_n la probabilidad de que el sistema vuelva por primera vez al estado inicial i, exactamente después de n pasos. El valor

$$v = \sum_{n=1}^{\infty} v_n$$

es la probabilidad de que el sistema caiga en el estado inicial i en un paso cualquiera $n = 1, 2, \ldots$, de otro modo, v es *la probabilidad de reversibilidad* a i.

El estado i se llama *reversible* si la probabilidad de volver a él es igual a 1, e *irreversible* si esta probabilidad es menor de 1.

Teorema 1. *El estado i es reversible, entonces y sólo entonces, cuando*

$$\sum_{n=1}^{\infty} p_{ii}(n) = \infty. \tag{2.7}$$

Demostración. Tiene lugar la igualdad siguiente:

$$u_n = u_0 v_n + u_1 v_{n-1} + \ldots + u_{n-1} v_1 + u_n v_0, \quad n = 1, 2, \ldots, \tag{2.8}$$

donde se supone que $u_n = p_{ii}(n)$ y se han introducido, complementariamente, $u_0 = 1$ y $v_0 = 0$. Esta es un corolario consecuencia de la fórmula general de la probabilidad total. Si introducimos, precisamente, los sucesos B_k, o sea, «el sistema vuelve al estado inicial i después de k pasos», $k = 1, \ldots, n$ y el suceso B_{n+1}, o sea, «el sistema no cae ni una vez en el estado i durante los n primeros pasos», entonces $B_1, \ldots, B_{n+1}$ será un sistema completo de sucesos, y la probabilidad del suceso A, o sea, «después de n pasos el sistema se encontrará en el estado inicial i», según la fórmula de la probabilidad total es

$$\mathbf{P}(A) = \sum_{i=1}^{n+1} \mathbf{P}(A \mid B_k) P(B_k),$$

donde

$$\mathbf{P}(B_k) = v_k;\ \mathbf{P}(A \mid B_k) = u_{n-k},\ k = 1, \ldots, n;\ P(A \mid B_{n+1}) = 0$$

Consideremos las llamadas *funciones productoras*

$$U(z)=\sum_{k=0}^{\infty} u_k z^k, \quad V(z)=\sum_{k=0}^{\infty} v_k z^k, \quad |z|\leqslant 1.$$

Evidentemente, para $|z|<1$ éstas son funciones analíticas. La relación (2.8), justa para todos los valores $n = 1, 2, \ldots$ se puede escribir en la forma

$$U(z) - u_0 = U(z)\, V(z), \quad u_0 = 1,$$

de donde

$$U(z)=\frac{1}{1-V(z)}. \tag{2.9}$$

La reversibilidad del estado i significa que

$$v=\sum_{k=0}^{\infty} v_k=\lim_{z\to 1} V(z)=1.$$

Como se ve de la igualdad (2.9) esta relación límite es equivalente a que

$$\lim_{z\to 1} U(z)=\lim_{z\to 1}\frac{1}{1-V(z)}=\infty.$$

Pero

$$\lim_{z\to 1} U(z)=\lim_{z\to 1}\sum_{n=0}^{\infty} u_n z^n=\sum_{n=0}^{\infty} u_n,$$

y de tal modo, la reversibilidad del estado i es equivalente a que la serie $\sum_{n=0}^{\infty} u_n$ sea divergente. Para terminar la demostración sólo queda por recordar que $u_n = p_{ii}(n)$.

Teorema 2. *Si el estado inicial i es reversible, el sistema vuelve a i un número infinito de veces, durante un número infinito de pasos con la probabilidad* 1. *Si este estado es irreversible, el sistema durante un número infinito de pasos estará en el estado i solamente un número finito de veces con la probabilidad* 1, *en otras palabras, después de cierto número finito de pasos el sistema no volverá nunca al estado i.*

Demostración. Designemos por ν_1 el número de pasos hasta el primer regreso al estado i, por ν_2 el número de pasos hasta el segundo regreso, etc. Si durante un número infinito de pasos se realizan menos de k regresos, entonces suponemos que $\nu_k = \infty$. El suceso $\{\nu_k < \infty\}$ significa que por lo menos se realizaron k regresos. La probabilidad del regreso es $\mathbf{P}\{\nu_1 < \infty\} = v$. En condiciones de realización del suceso $\{\nu_1 < \infty\}$, el sistema después de un número finito ν_1 de pasos regresa al estado inicial i, después de lo cual su futuro comportamiento se somete a las mismas leyes, como si él sólo

comenzara su movimiento. De tal forma, la probabilidad del suceso $\{\nu_2 < \infty\}$ en condiciones de que $\{\nu_1 < \infty\}$, también será igual a v:

$$\mathbf{P}\{\nu_2 < \infty \mid \nu_1 < \infty\} = v,$$

Evidentemente, si $\nu_1 = \infty$, también será $\nu_2 = \infty$. Por eso

$$\mathbf{P}\{\nu_2 < \infty\} = \mathbf{P}\{\nu_2 < \infty \mid \nu_1 < \infty\} \cdot \mathbf{P}\{\nu_1 < \infty\} = v^2.$$

De modo absolutamente análogo, para cualquier k,

$$\mathbf{P}\{\nu_k < \infty \mid \nu_{k-1} < \infty\} = v, \quad \mathbf{P}\{\nu_k < \infty\} = v^k.$$

La irreversibilidad del estado i significa, que $v < 1$. En este caso

$$\sum_{k=1}^{\infty} \mathbf{P}\{\nu_k < \infty\} = \sum_{k=1}^{\infty} v^k < \infty$$

y, de acuerdo al lema de Borel—Cantelli (véase la pag. 63), puede ocurrir sólo un número finito de sucesos del tipo $\{\nu_k < \infty\}$ con la probabilidad 1; es decir, durante un número infinito de pasos el sistema puede encontrarse en el estado i con la probabilidad 1, sólo un número finito de veces.

La reversibilidad del estado i significa que $v = 1$. En este caso, para cualquier k

$$\mathbf{P}\{\nu_k < \infty\} = 1.$$

Designemos por $\varkappa$ al número de regresos durante un número infinito de pasos. Evidentemente, el suceso $\{\varkappa \geqslant k\}$ es idéntico al suceso $\{\nu_k < \infty\}$, y $\{\varkappa = \infty\} = \bigcap_k \{\nu_k < \infty\}$. Debido a la continuidad de la probabilidad

$$\mathbf{P}\{\varkappa = \infty\} = \lim_{k \to \infty} \mathbf{P}\{\nu_k < \infty\} = 1.$$

El teorema queda demostrado.

Ejemplo (fluctuación aleatoria). Examinemos la fluctuación aleatoria de una partícula por los puntos de números enteros de la recta real, por la cual la partícula se desplaza a 1 en dirección positiva con probabilidad p y en dirección negativa con probabilidad $q = (1 - p)$. Según la fórmula (1.1)

$$p_{ii}(2n) = \frac{(2n)!}{n!\,n!}[pq]^n \sim \frac{[4pq]^n}{\sqrt{\pi n}},$$

y en la fluctuación aleatoria simétrica, cuando $p = q = 1/2$, $p_{ii}(2n) \sim 1/\sqrt{\pi n}$ y la serie $\sum_n p_{ii}(2n)$ es divergente, todos los estados son reversibles. Para $p \neq q$, cuando $4pq < 1$ y $\sum_n p_{ii}(2n) < \infty$, todos los estados son irreversibles.

Hallemos la probabilidad del regreso. Supongamos que en el punto 0 se tiene la tal llamada pantalla absorbente: al caer la partícula en el punto 0 se queda allí para siempre. Evidentemente, desde cualquier estado $i > 0$ se puede caer en cualquier estado $j > 0$ con probabilidad positiva. Designemos por v_{ij} la probabilidad de caer alguna vez desde i en j. Aquí tiene lugar la siguiente relación:

$$v_{ij} = pv_{i+1,\,j} + qv_{i-1,\,j},$$

que nos da la fórmula de la probabilidad total. Esta relación representa en sí una ecuación en diferencias finitas para v_{ij} como función de $i = 1, 2, \ldots$. Con objeto de determinar la solución v_{ij} de esta ecuación, son necesarias además las «condiciones límites». Ellas se pueden hallar de las siguientes consideraciones. Las probabilidades v_{ij} para $i \neq j$ no se cambian si en el punto j ponemos también una pantalla absorbente. En este caso, evidentemente, tendremos $v_{0j} = 0$ y $v_{jj} = 1$. Respondiendo a estas «condiciones límites» la función v_{ij} de $i = 1, 2, \ldots, j - 1$ para $p \neq q$ tiene la forma

$$v_{ij} = A_j + B_j \left(\frac{q}{p}\right)^i, \quad 0 < i < j,$$

donde

$$A_j = \frac{1}{1 - \left(\frac{q}{p}\right)^j}, \quad B_j = -\frac{1}{1 - \left(\frac{q}{p}\right)^j};$$

si es $p = q = 1/2$, entonces

$$v_{ij} = \frac{i}{j}, \quad 0 < i < j.$$

Está claro que tienen lugar fórmulas análogas cuando la pantalla absorbente está en un punto determinado k y $k < i < j$ (en las partes derechas de las fórmulas indicadas sólo es necesario sustituir i por $i - k$ y j por $j - k$). Si $k \to -\infty$, entonces la influencia de la pantalla absorbente desaparece y las fórmulas límites dan la expresión para la probabilidad v_{ij} de paso del punto i al punto $j > i$ en la fluctuación aleatoria corriente sin cualquier pantalla absorbente. Estas fórmulas tienen la forma

$$v_{ij} = \begin{cases} \left(\frac{p}{q}\right)^{j-1} & \text{para } p < 1/2, \\ 1 & \text{para } p \geqslant 1/2. \end{cases} \quad (i \leqslant j).$$

Sustituyendo aquí p por q, obtenemos las expresiones de las probabilidades v_{ij} para $i > j$, a saber:

$$v_{ij} = \begin{cases} \left(\frac{q}{p}\right)^{i-j} & \text{para } p > 1/2, \\ 1 & \text{para } p \leqslant 1/2. \end{cases} \quad (i \geqslant j).$$

Todas estas fórmulas han sido obtenidas con la suposición de que en el punto j está la pantalla absorbente, y naturalmente, para $i = j$, dan el valor $v_{ij} = 1$. La verdadera probabilidad v de regreso al estado inicial $i = j$ puede ser determinada según las probabilidades ya encontradas v_{ij} para $i \neq j$, como

$$v = pv_{i+1,\,i} + qv_{i-1,\,i},$$

lo que da la siguiente expresión para la probabilidad de regreso:

$$v = \begin{cases} 2p & \text{para } p < 1/2 \\ 2q & \text{para } q > 1/2 \\ 1 & \text{para } p = q \end{cases} \Bigg\} = 1 - |p - q|.$$

3. Tiempo medio de estancia en un estado. Clasificación de los estados. Volvamos a la cadena general de Markov $\xi(n)$, $n = 0, 1, \ldots$. Supongamos que en el momento inicial de tiempo el sistema se encuentra en el estado $\xi(0) = i$. Designemos por ν al número de pasos hasta el primer regreso a este estado (si consideramos que cada paso ocupa una unidad de tiempo, entonces ν será el tiempo de regreso a i). Para el estado reversible el tiempo de regreso con probabilidad 1 es naturalmente $\mathbf{P}\{\nu < \infty\} = 1$; para el estado irreversible

$$\mathbf{P}\{\nu < \infty\} = v < 1, \quad \mathbf{P}\{\nu = \infty\} = 1 - v > 0. \tag{2.10}$$

Introduzcamos el valor medio

$$\mu = \mathbf{M}\nu = \sum_{n=1}^{\infty} n\mathbf{P}\{\nu = n\}, \tag{2.11}$$

suponiendo naturalmente, que $\mu = \infty$ para el estado reversible i.

Examinemos al principio el caso cuando $\mu < \infty$. De acuerdo con el teorema 2, en el transcurso del tiempo el sistema regresará de nuevo y de nuevo al estado inicial i. Designemos por ν_1 $(=\nu)$ el momento del primer regreso, por ν_2 el momento del segundo regreso . . ., por ν_n el momento del regreso n al estado i. Como ya indicamos en la demostración del teorema 2, el comportamiento del sistema después del regreso, en uno de los momentos ν_k, al estado i, no depende (para $\xi(\nu_k) = i$) de su comportamiento hasta el momento ν_k, y se somete a las mismas leyes que rigen desde el principio, para el estado inicial $\xi(0) = i$. En particular, las magnitudes $\tau_1 = \nu_1$, o sea, el tiempo del primer regreso, $\tau_2 = (\nu_2 - \nu_1)$, el tiempo del segundo regreso . . ., $\tau_n = (\nu_n - \nu_{n-1})$, el tiempo del n-ésimo regreso al estado inicial i serán magnitudes aleatorias independientes con la misma distribución de probabilidades, que tienen el valor medio μ. Según la ley refor-

zada de los grandes números (véase (5.7), cap. I)),

$$\frac{1}{n}\sum_{k=1}^{n}\tau_k \to \mu \quad \text{para } n \to \infty \tag{2.12}$$

con probabilidad 1.

En esta relación $\nu_n = \sum_{k=1}^{n}\tau_k$ es el tiempo durante el cual se realizaron exactamente n regresos al estado i (recordemos que según las condiciones cada paso se prolonga en una unidad de tiempo), en otras palabras, n es el tiempo total que el sistema empleó en el estado i desde el intervalo de tiempo 1 hasta el ν_n. Designemos por n_N el tiempo total empleado por el sistema en el estado i durante el intervalo arbitrario desde 1 hasta N ($n = n_N$ es el número de caídas en el estado i durante el tiempo desde 1 hasta N). Ya que $n = n_N \to \infty$ para $N \to \infty$ (véase el teorema 2), entonces de la relación límite (2.12) se deduce que

$$\frac{\nu_n}{n} \to \mu, \quad \frac{n}{n+1} \to 1, \quad \frac{\nu_{n+1}}{n} \to \mu \text{ para } N \to \infty$$

con la probabilidad 1. Evidentemente, debido a la propia definición del número n_N, tenemos $\nu_{n_N} \leqslant N \leqslant \nu_{n_N+1}$ y, por consiguiente,

$$\frac{\nu_N}{n_N} \to \mu \quad \text{para } N \to \infty \tag{2.13}$$

con probabilidad 1.

Mostremos ahora que la relación límite (2.13) tiene lugar también en el caso, cuando el valor medio μ, determinado por la igualdad (2.11), es infinito ($\mu = +\infty$). Para el estado irreversible i esto se deduce inmediatamente de la segunda parte del teorema 2 (n_N queda limitado para $N \to \infty$ con probabilidad 1). Para el estado reversible i nos dirigimos de nuevo a las magnitudes independientes $\tau_k = \nu_k - \nu_{k-1}$, $k = 1, 2, \ldots$. Determinemos las magnitudes «cortadas» τ_k^m ($\tau_k^m \leqslant \tau_k$), poniendo

$$\tau_k^m = \begin{cases} \tau_k & \text{para } \tau_k \leqslant m, \\ 0 & \text{para } \tau_k > m. \end{cases}$$

El valor medio

$$\mu^m = \mathbf{M}\tau_k^m = \sum_{n=1}^{m} n\mathbf{P}(\tau_k^m = n)$$

de estas magnitudes (véase (2.11)) para los valores de m suficientemente grandes, se puede hacer tan grande como se quiera, ya que $\sum_{n=1}^{\infty} n\mathbf{P}(\tau_k = n) = \infty$. Según la ley reforzada de los grandes números,

$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{m} \tau_k^m = \mu^m$ con probabilidad 1, y por eso

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \tau_k \geqslant \lim_{n\to\infty} \sum_{k=1}^{n} \tau_k^m = \mu^m.$$

De tal modo, el límite inferior $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \tau_k$ supera al número tan grande como se quiera μ^m, es decir

$$\frac{1}{n} \sum_{k=1}^{n} \tau_k \to \infty \text{ para } n \to \infty \tag{2.14}$$

con probabilidad 1. Exactamente igual que de (2.12), de la relación (2.14) se deduce (2.13) con $\mu = \infty$.

La relación n_N/N entre las magnitudes n_N y N que figuran en la relación límite (2.13) representa en sí la parte de tiempo, empleado por el sistema en el estado i durante el intervalo de tiempo desde 1 hasta N, y (2.13) es equivalente a la siguiente proposición:

Teorema 3. *Con la probabilidad* 1

$$\frac{n_N}{N} \to \frac{1}{\mu} \quad \textit{para} \quad N \to \infty \tag{2.15}$$

donde μ *es el tiempo medio de regreso al estado* i.

Se comprende fácilmente que este resultado queda en vigor para cualquier estado i (que no es obligatoriamente inicial) con tal de que el sistema caiga tarde o pronto en este estado (después de lo cual su comportamiento se somete a las mismas leyes que rigen el estado inicial i).

De la relación (2.15) se deduce (véase el teorema 5 del § 4, cap. I) que $1/\mu$ coincide con el límite $\lim_{N\to\infty} \mathbf{M}\left[\frac{n_N}{N}\right]$ de las esperanzas matemáticas de las magnitudes aleatorias limitadas n_N/N. Con objeto de hallar la expresión explícita de la esperanza matemática $\mathbf{M}n_N$ del número de caídas en el estado i durante el tiempo N, introducimos las magnitudes

$$\chi_k = \begin{cases} 1 & \text{para } \xi(k) = i, \\ 0 & \text{en los casos restantes.} \end{cases}$$

Evidentemente,

$$n_N = \sum_{k=1}^{N} \chi_k, \quad \mathbf{M}n_N = \sum_{k=1}^{N} \mathbf{M}\chi_k = \sum_{k=1}^{N} p_{ii}(k). \tag{2.16}$$

En resultado obtenemos que

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} p_{ii}(k) = \frac{1}{\mu} . \qquad (2.17)$$

El estado i se llama *positivo* cuando $\frac{1}{\mu} > 0$, y nulo cuando $\frac{1}{\mu} = 0$ (recordemos que de acuerdo con (2.15) la magnitud $1/\mu$ es la parte de tiempo, empleado por el sistema en el estado i en un intervalo infinito de tiempo).

Se dice que el estado j es *accesible* desde i, si durante un determinado número de pasos el sistema pasa desde i hasta j con probabilidad positiva, es decir, $p_{ij}(M) > 0$ para un M dado. Si j es accesible desde i y a su vez i es accesible desde j, entonces estos estados i, j se llaman *comunicantes*.

Sea i un estado reversible y j accesible desde i. Entonces i es a la vez accesible desde j, ya que en caso contrario el sistema saliendo, desde i, después de M pasos, cae en el estado j con probabilidad positiva $p_{ij}(M) = \alpha > 0$, después de lo cual ya no puede regresar a i; de tal modo, la probabilidad del regreso a i no será mayor que $1 - \alpha$, y esto contradice a la reversibilidad del estado i. Así pues, si j es accesible del estado reversible i, entonces a su vez, i es accesible desde j, es decir $p_{ij}(N) = \beta > 0$ para un N dado, es decir, *los estados i y j son comunicantes*.

Para cualesquiera estados comunicantes i, j (tales que $p_{ij}(M) = \alpha > 0$, $p_{ji}(N) = \beta > 0$ para unos determinados M y N) de la relación

$$P(M + K + N) = P(M)\, P(K)\, P(N) = P(N)\, P(K)\, P(M)$$

(donde $P(n) = \{p_{kl}(n)\}$ significa la matriz de la probabilidad de paso $p_{kl}(n) \geqslant 0$, véase (2.2) y más adelante) deducimos que

$$p_{ii}(K + M + N) \geqslant p_{ij}(M)\, p_{jj}(K)\, p_{ji}(N) = \alpha\beta p_{jj}(K),$$
$$p_{jj}(K + M + N) \geqslant p_{ji}(N)\, p_{ii}(K)\, p_{ij}(M) = \alpha\beta p_{ii}(K).$$

Estas desigualdades muestran, primeramente que las series

$$\sum_k p_{ii}(k) \quad \sum_k p_{jj}(k)$$

convergen o divergen al mismo tiempo y segundo, que los límites

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} p_{ii}(k), \quad \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} p_{jj}(k)$$

bien ambos son iguales a cero, o bien ambos son positivos. Teniendo en cuenta el teorema 2, de aquí sacamos la conclusión de que *los estados accesibles desde cierto estado reversible, también son reversibles.*

Evidentemente (véanse (2.7) y (2.17)), los estados comunicantes pertenecen a un mismo tipo: son reversibles o irreversibles, positivos o nulos.

Sea i un estado reversible. Como fue indicado, todos los estados j accesibles desde i son reversibles y se comunican con el estado i. Todos los estados indicados j forman la llamada clase *cerrada* de estados (designémosla por E) tal que si el sistema en un paso cae en uno de los estados $j \in E$, entonces, después de esto, se queda para siempre en el conjunto de estados E (cualquier estado, accesible desde cualquier estado $j \in E$, también entra en E); con ello, para cualquier $j \in E$, se puede determinar la propia clase como un conjunto de todos los estados accesibles desde j.

Examinemos dos clases cerradas cualesquiera de estados reversibles designándolas por E_1 y E_2. Supongamos que un estado j entra simultáneamente en E_1 y en E_2. Entonces, estas clases E_1 y E_2 coinciden, ya que cada una de ellas coincide con el conjunto de todos los estados, accesibles desde j. Por consiguiente, bien coinciden las clases cerradas E_1 y E_2, o bien no se cortan (no tienen estados comunes j).

Separemos el conjunto E_0 de todos los estados irreversibles y dividamos los estados reversibles restantes en clases cerradas que no se cortan E_1, E_2, . . . (cada una de las clases representa en sí un conjunto determinado de estados reversibles que se comunican uno con otro).

Como muestra el ejemplo de la fluctuación aleatoria en la pág. 147, hablando en general, el sistema en el transcurso del tiempo puede marchar «al infinito» por una cadena de estados irreversibles en E_0 (si hay un número infinito de tales estados). Esta posibilidad se excluye, si sólo se tiene un número finito de estados irreversibles. En este caso, el sistema estará en cualesquiera de los estados irreversibles, sólo un número finito de veces con probabilidad 1 y tarde o temprano caerá en un determinado estado reversible i; en el futuro «circulará» en la clase cerrada de estados reversibles E, que contiene i (así, precisamente, ocurre con la cadena con un número finito de estados).

4. Teorema ergódico (convergencia hacia la distribución estacionaria). La distribución de probabilidades p_i^*, $i = 1, 2, \ldots$, se llama distribución *estacionaria* para las cadenas de Markov con probabilidades de paso p_{ij} si se cumple la condición

$$p_j^* = \sum_i p_i^* p_{ij} \quad j = 1, 2, \ldots \tag{2.18}$$

De las relaciones recurrentes generales (2.1) se ve que la condición (2.18) significa lo siguiente: *en la distribución* $p_i(n_0) = p_i^*$, $i = 1, 2, \ldots$, *las probabilidades* $p_j(n)$ *de que en el paso* n *el sistema se encuentre en el estado* $j = 1, 2, \ldots$, *quedan invariables para todos los valores* $n \geqslant n_0$:

$$p_j(n) = p_j^*, \quad j = 1, 2, \ldots \tag{2.19}$$

Supongamos que los estados de la cadena de Markov examinada son positivos, forman una clase cerrada y además de esto, para cierto n_0, es positivo *el coeficiente de ergodicidad* $\boldsymbol{k}(n_0)$ que se determina por la fórmula

$$\boldsymbol{k}(n_0) = 1 - \frac{1}{2} \sup_{i,j} \sum_m | p_{im}(n_0) - p_{jm}(n_0) |.$$

Señalemos que para el coeficiente de ergodicidad $\boldsymbol{k}(n_0)$ tiene lugar la siguiente valuación sencilla

$$\boldsymbol{k}(n_0) \geqslant \sup_j \inf_i p_{ij}(n_0), \tag{2.20}$$

de la cual se deduce que $\boldsymbol{k}(n_0) > 0$, si existe aunque sea un sólo estado j, accesible desde cualquier estado i (para un mismo número de pasos n_0) con probabilidad $p_{ij}(n_0)$, $p_{ij}(n_0) \geqslant \delta > 0$ para todos los valores $i = 1, 2, \ldots$

Señalemos también que la condición $\boldsymbol{k}(n_0) > 0$ no siempre se cumple. De ejemplo trivial puede servir la cadena de Markov con dos estados 1 y 2, cuando el sistema cambia periódicamente (a cada paso) su estado (en este caso $p_{11}(n) = 1$, $p_{21}(n) = 0$, $p_{12}(n) = 0$, $p_{22}(n) = 1$ para los valores pares de n, $p_{11}(n) = 0$, $p_{21}(n) = 1$, $p_{12}(n) = 1$, $p_{22}(n) = 0$ para los n impares y $\sum_k | p_{1k}(n) - p_{2k}(n) | \equiv 2$ para todos los n).

Teorema 4. Si $\boldsymbol{k}(n_0) > 0$ para un determinado n_0, entonces existen los límites

$$\lim_{n \to \infty} p_j(n) = p_j^*, \quad j = 1, 2, \ldots, \tag{2.21}$$

donde las probabilidades p_j^*, $j = 1, 2, \ldots$, *forman la distribución estacionaria. Para cualquier distribución inicial* p_j^0, $i = 1, 2, \ldots$, *las probabilidades límites* p_j^*, $j = 1, 2, \ldots$, *son las mismas, con ello tiene lugar la siguiente valuación de la velocidad de convergencia en la relación límite* (2.21):

$$\sup_{p_i^0} | p_j(n) - p_j^* | \leqslant Ce^{-Dn} \tag{2.22}$$

$$\left(C = \frac{1}{1 - \boldsymbol{k}(n_0)}, \; D = \frac{1}{n_0} \ln \frac{1}{1 - \boldsymbol{k}(n_0)} \right).$$

Demostración. Admitamos que

$$r_j(n) = \inf_i p_{ij}(n), \quad R_j(n) = \sup_i p_{ij}(n).$$

Tenemos

$$r_j(n+1) = \inf_i p_{ij}(n+1) = \inf_i \sum_k p_{ik} p_{kj}(n) \geqslant \inf_i \sum_k p_{ik} r_j(n) = r_j(n),$$

$$R_j(n+1) = \sup_i p_{ij}(n+1) = \sup_i \sum_k p_{ik} p_{kj}(n) \leqslant \sup_i \sum_k p_{ik} R_j(n) = R_j(n).$$

De tal modo, obtenemos la siguiente cadena de desigualdades:

$$r_j(1) \leqslant r_j(2) \leqslant \ldots \leqslant r_j(n) \leqslant \ldots \leqslant R_j(n) \leqslant \ldots \leqslant R_j(2) \leqslant R_j(1).$$

Para cualesquiera estados α y β

$$\sum_k p_{\alpha k}(n_0) - \sum_k p_{\beta k}(n_0) =$$

$$= \sum_k^+ [p_{\alpha k}(n_0) - p_{\beta k}(n_0)] + \sum_k^- [p_{\alpha k}(n_0) - p_{\beta k}(n_0)] = 0,$$

donde $\sum_k^+$ significa la adición por todos los valores de k, para los cuales la diferencia $p_{\alpha k}(n_0) - p_{\beta k}(n_0)$ es positiva, y $\sum_k^-$ es la adición por aquellos valores de k, para los cuales la diferencia $p_{\alpha k}(n_0) - p_{\beta k}(n_0)$ es negativa. Evidentemente,

$$\sup_{\alpha, \beta} \sum_k^+ [p_{\alpha k}(n_0) - p_{\beta k}(n_0)] = 1 - \boldsymbol{k}(n_0),$$

donde $\boldsymbol{k}(n_0)$ es el coeficiente de ergodicidad. Utilizando las relaciones obtenidas, valoremos la diferencia $R_j(N) - r_j(N)$. Tenemos

$$R_j(n_0) - r_j(n_0) = \sup_\alpha p_{\alpha j}(n_0) - \inf_\beta p_{\beta j}(n_0) =$$

$$= \sup_{\alpha, \beta} [p_{\alpha j}(n_0) - p_{\beta j}(n_0)] \leqslant \sup_{\alpha, \beta} \sum_k^+ [p_{\alpha k}(n_0) - p_{\beta k}(n_0)] = [1 - \boldsymbol{k}(n_0)]$$

y

$$R_j(n_0 + n) - r_j(n_0 + n) = \sup_{\alpha, \beta} [p_{\alpha j}(n_0 + n) - p_{\beta j}(n_0 + n)] =$$

$$= \sup_{\alpha, \beta} \sum_k [p_{\alpha k}(n_0) - p_{\beta k}(n_0)]\, p_{kj}(n) \leqslant$$

$$\leqslant \sup_{\alpha, \beta} \left\{ \sum_k^+ [p_{\alpha k}(n_0) - p_{\beta k}(n_0)]\, R_j(n) + \sum_k^- [p_{\alpha k}(n_0) - p_{\beta k}(n_0)]\, r_j(n) \right\} =$$

$$= \sup_{\alpha, \beta} \left\{ \sum_k^+ [p_{\alpha k}(n_0) - p_{\beta k}(n_0)] \, [R_j(n) - r_j(n)] \right\} =$$

$$= [1 - \boldsymbol{k}(n_0)]\, [R_j(n) - r_j(n)].$$

De aquí se deduce que

$$R_j(Nn_0) - r_j(Nn_0) \leqslant [1 - \boldsymbol{k}(n)]^N, \quad N = 1, 2, \ldots$$

La sucesión $r_j(n)$, $n = 1, 2, \ldots$, crece monótonamente y la sucesión $R_j(n)$, $n = 1, 2 \ldots$ decrece monótonamente, siendo $r_j(n) \leqslant R_j(n)$. La valuación, obtenida anteriormente, de la diferencia $R_j(n) - r_j(n)$ muestra que estas sucesiones tienen un mismo límite p_j^*:

$$p_j^* = \lim_{N \to \infty} r_j(n) = \lim_{N \to \infty} R_j(n).$$

Evidentemente

$$|\, p_{ij}(n) - p_j^* \,| \leqslant R_j(n) - r_j(n) \leqslant (1 - k(n_0))^{\frac{n}{n_0}-1}.$$

Para cualquier distribución inicial p_j^0, $j = 1, 2, \ldots$, tenemos

$$p_j(n) - p_j^* = \Big|\sum_i p_i^0 p_{ij}(n) - p_j^*\Big| \leqslant \sum_i p_i^0 \,|\, p_{ij}(n) - p_j^* \,| \leqslant$$

$$\leqslant \sum_i p_i^0 [R_j(n) - r_j(n)] = R_j(n) - r_j(n) \leqslant (1 - k(n_0))^{\frac{n}{n_0}-1}.$$

La valuación obtenida se puede escribir, naturalmente, en la forma (2.22) con las respectivas constantes C y D, suponiendo

$$C = \frac{1}{1 - k(n_0)}, \quad D = \frac{1}{n_0} \ln \frac{1}{1 - k(n_0)}.$$

Mostremos que *las probabilidades límites p_j^* satisfacen al sistema de ecuaciones* (2.18). Para esto, señalemos al principio que la suma $\sum_j p_j^*$ es finita, ya que para cualquier m

$$\sum_{j \leqslant m} p_j^* = \lim_{n \to \infty} \sum_{j \leqslant m} p_j(n) \leqslant 1.$$

Luego, de la desigualdad $p_j(n+1) \geqslant \sum_{i \leqslant m} p_i(n)\, p_{ij}$ (véase (2.1)) deducimos que $p_j^* \geqslant \sum_{i \leqslant m} p_i^* p_{ij}$ para cualquier valor limitado de m, y, por consiguiente,

$$p_j^* \geqslant \sum_i p_i^* p_{ij}, \quad j = 1, 2, \ldots \tag{2.23}$$

Sumando todas estas desigualdades, obtenemos

$$\sum_j p_j^* \geqslant \sum_j \Big(\sum_i p_i^* p_{ij}\Big) = \sum_i p_i^* \Big(\sum_j p_{ij}\Big) = \sum_i p_i^*.$$

Aquí se ve que debe existir la igualdad rigurosa $\Big(\sum_j p_j^* = \sum_j p_j^*\Big)$, y por consiguiente, deberá existir la igualdad rigurosa en todas las relaciones (2.23), que es lo que se exigía demostrar.

Establezcamos luego que todas las probabilidades p_j^* son positivas y $\sum_j p_j^* = 1$. Evidentemente,

$$p_j^* = \lim_{n \to \infty} p_{ij}(n) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}(n),$$

y, por consiguiente, las probabilidades finales p_j^* para todos los estados positivos serán tales que

$$p_j^* = \frac{1}{\mu_j} > 0, \quad j = 1, 2, \ldots, \tag{2.24}$$

donde μ_j es el tiempo medio de regreso al estado j (véase (2.17)). Luego, si tomamos en calidad de distribución inicial de probabilidades a

$$p_j^0 = p_i^* \frac{1}{\sum\limits_j p_j^*}, \quad i = 1, 2, \ldots,$$

entonces, ya que las probabilidades p_i^0 se diferencian de p_i^* ($i = 1, 2, \ldots$) sólo en un factor constante $(\sum\limits_j p_j^*)^{-1}$, éstas satisfacerán al sistema homogéneo de ecuaciones (2.18)

$$p_j^0 = \sum_i p_i^0 p_{ij}, \quad j = 1, 2, \ldots$$

Para la distribución estacionaria p_i^0, $i = 1, 2, \ldots$ tenemos

$$p_i(n) = p_j^0 \quad \text{y} \quad p_j^* = \lim_{n\to\infty} p_j(n) = p_j^0, \quad j = 1, 2, \ldots,$$

de donde se desprende que $\sum\limits_j p_j^* = 1$. De tal modo, las probabilidades límites p_i^*, $i = 1, 2, \ldots$, forman la distribución de probabilidades estacionaria (que evidentemente es única).

Señalemos que las probabilidades límites p_j^*, como cualesquiera probabilidades estacionarias, pueden ser halladas del sistema de ecuaciones lineales (2.18).

El teorema queda demostrado.

Ejemplo (*pila de libros*). Sobre la mesa del escritorio descansa una pila de m libros. Si designamos a cada libro con un número correspondiente, entonces el orden de su disposición desde arriba abajo se puede describir por la permutación de m números $(i_1, i_2, \ldots, i_m)$, donde i_1 es el número del libro que descansa arriba, i_2 es el número del libro siguiente, $\ldots$, i_m es el número del último libro que descansa lo más abajo. Supongamos que cada libro se coge con una probabilidad determinada, digamos, el libro con número k se coge con la probabilidad p_k ($k = 1, \ldots, m$), con ello al devolverlo se coloca encima.

Tomemos un estado arbitrario $(i_1, i_2, \ldots, i_m)$. En el siguiente paso aquél se encuentra o bien invariable, lo que ocurre con probabilidad p_i, al elegir el libro que descansa desde arriba con el número i_1, o bien se cambia por uno de los $m - 1$ estados de la forma $(i_k, i_1 \ldots)$, lo que ocurre con probabilidad p_{i_k} al elegir el libro con el número $i_k \neq i_1$.

Ante nosotros tenemos la cadena de Markov con los estados, cada uno de los cuales se describe por la correspondiente permutación $(i_1, i_2, \ldots, i_m)$ y por las probabilidades de paso indicadas. Si un libro determinado i no se elige nunca ($p_i = 0$), entonces todos los estados $(i_1, \ldots, i_m)$, donde $i_1 = i$ (el libro con el número i descansa

arriba), son irreversibles, ya que después del primer paso se elige un libro determinado j, $j \neq i$ y en lo sucesivo el i-ésimo libro que no se elige nunca, desciende hacia abajo. Si cada libro se elige con probabilidad positiva $p_l > 0$, entonces cada estado es accesible desde cualquier otro estado (en total se tienen $m!$ estados diferentes, o sea, las distintas permutaciones $(i_1, \ldots, i_m)$), y todos estos estados comunicantes son positivos, formando una clase cerrada. De cada estado $(i_1, \ldots, i_m)$ se puede pasar después de m pasos a cualquier estado $(j_1, \ldots, j_m)$ con probabilidad superior al producto $p_i \ldots p_m$ (este producto es igual a la probabilidad de paso de $(i_1, \ldots, i_m)$ a $(j_1, \ldots \ldots, j_m)$, cuando se elige en el primer paso el libro con el número j_m, en el segundo paso, el libro con el número j_{m-1} y así sucesivamente, en el último paso m se elige el libro con el número j_1). Por consiguiente, se tiene un coeficiente de ergodicidad positivo $k(m)$ ($k(m) \geqslant p_1$, $p_2, \ldots, p_m$; véase (2.20)) y en el transcurso del tiempo se establece la distribución de probabilidades estacionaria.

Examinemos al principio el caso $m = 2$. Entonces se tienen sólo dos estados (1,2) y (2,1). Las probabilidades de paso tienen la forma

$$p_{11} = p_{21} = p_1, \quad p_{12} = p_{22} = p_2,$$

y la matriz de las probabilidades de paso es

$$P = \begin{pmatrix} p_1 & p_2 \\ p_1 & p_2 \end{pmatrix}.$$

Las probabilidades de paso en dos pasos son

$$p_{11}(2) = p_{21}(2) = p_1 p_1 + p_2 p_1 = p_1 (p_1 + p_2) = p_1,$$
$$p_{12}(2) = p_{22}(2) = p_1 p_2 + p_2 p_2 = p_2 (p_1 + p_2) = p_2.$$

Se ve que $P^2 = P$ y en general, $P^n = P$. Para cualquier distribución inicial de probabilidades tenemos

$$p_1(n) = p_1^0 p_{11}(n) + p_2^0 p_{21}(n) = p_1 (p_1^0 + p_2^0) = p_1,$$
$$p_2(n) = p_1^0 p_{12}(n) + p_2^0 p_{22}(n) = p_2 (p_1^0 + p_2^0) = p_2.$$

Se ve que la distribución estacionaria se establece ya en el primer paso.

Examinemos el caso del valor arbitrario de m. Designemos por $p_{(i_1, \ldots, i_m), (j_1, \ldots, j_m)}$ la probabilidad del paso del estado $(i_1, \ldots, i_m)$ al estado $(j_1, \ldots, j_m)$. Como fue mostrado,

$$p_{(i_1, \ldots, i_m), (j_1, \ldots, j_m)} = \begin{cases} p_{i_k} \text{ para } (j_1, \ldots, j_m) = (i_k, \ldots), \\ 0 \quad \text{para los restantes } (j_1, \ldots, j_m), \end{cases}$$

donde la permutación $(i_k, \ldots)$ se obtiene de $(i_1, \ldots, i_m)$ con la elección de un determinado i_k y su permutación en el primer lugar. Las probabilidades estacionarias $p^*_{(j_1, \ldots, j_m)}$ son la solución del

siguiente sistema de ecuaciones lineales (véase (2.18)):

$$p^*_{(i_1, \ldots, i_m)} = p_{i_1} \sum_{k=1}^{m} p^*_{(i_2, \ldots, i_{k-1}, i_1, i_k, \ldots)}.$$

Es natural preguntar, ¿con qué probabilidad cada uno de los libros existentes resultará estar colocado arriba? ¿Cuándo se establecerá prácticamente, después de un número bastante grande de pasos, la distribución de probabilidades estacionaria (es decir, cuándo ocupará la pila de libros la posición correspondiente $(i_1, \ldots, i_m)$ con unas probabilidades invariables $p^*_{(i_1, \ldots, i_m)}$?

La probabilidad de que el libro con el número i descanse arriba, será evidentemente,

$$p_i^* = \sum_{i_2, \ldots, i_m} p^*_{(i, i_2, \ldots, i_m)},$$

donde la suma se toma según todos los estados, en los cuales está en primer lugar i. De las ecuaciones para las probabilidades estacionarias obtenemos que

$$p_i^* = \sum_{i_2, \ldots, i_m} p_i \sum_k p^*_{(i_2, \ldots, i_{k-1}, i, i_k, \ldots, i_m)} =$$

$$= p_i \sum_{i_1, \ldots, i_m} p^*_{(i_1, \ldots, i_m)} = p_i, \quad i = 1, \ldots, m,$$

es decir, la probabilidad p_i^* de que el libro con el número i se encuentre arriba, será igual a la probabilidad p_i, con la cual se elige este libro. De tal modo, cuanto más frecuentemente se tome tal o cual libro, tanto mayor será la probabilidad de que éste se encuentre arriba.

Ejemplo (*la fluctuación aleatoria*). Examinemos la fluctuación aleatoria, en la cual la partícula con probabilidad p_i pasa del punto i al punto vecino $j = i + 1$, y con probabilidad $q_i = 1 - p_i$, al punto $j = 0$. Consideraremos que $0 < p_i < 1$. Evidentemente, todos los estados son comunicantes y al mismo tiempo son reversibles o irreversibles, nulos o positivos.

Supongamos que la partícula se encuentra en el estado inicial $i = 0$. La probabilidad de que después de los n pasos siguientes no regrese ni una vez a la posición inicial $i = 0$, será igual al producto $p_0 p_1 \ldots p_{n-1}$, o sea, la probabilidad de que el sistema recorra sucesivamente la cadena de estados $0 \to 1 \to \ldots \to n$. Se ve fácilmente, que la probabilidad de que éste no regrese ni una vez al estado inicial $i = 0$, después de un número infinito de pasos, será igual al producto infinito

$$\prod_{k=0}^{\infty} p_k = \lim_{n \to \infty} p_0 p_1 \ldots p_n.$$

Si este producto infinito es convergente hacia cero: $\lim_{n\to\infty} p_0 p_1 \ldots p_n = 0$, entonces el estado $i = 0$ es reversible. En caso contrario la probabilidad de regreso es

$$v = 1 - \lim_{n\to\infty} p_0 p_1 \ldots p_n < 1,$$

y el estado $i = 0$ es irreversible.

A este resultado se puede llegar por otro camino, examinando directamente las probabilidades v_n de regresar por primera vez al *estado inicial 0, exactamente después de n pasos*. Evidentemente, la partícula se vuelve por primera vez al estado 0 en el paso n, si ella en los primeros $n - 1$ pasos pasa, sucesivamente, desde el estado $i - 1$ al i (con probabilidades p_{i-1}, $i = 1, \ldots, n-1$), y por eso $v_1 = 1 - p_0$, $v_n = p_0 \ldots p_{n-2}(1 - p_{n-1})$, $n = 2, 3, \ldots$ La probabilidad de regreso al estado inicial 0, según la definición, será igual a la suma $\sum_{n=1}^{\infty} v_n$ y es

$$v = \sum_{n=1}^{\infty} v_n = 1 - p_0 + [p_0(1 - p_1)] + \ldots = 1 - \lim_{n\to\infty} p_0 p_1 \ldots p_n.$$

Para los estados irreversibles, cuando $v < 1$, la partícula se alejará en dirección $+\infty$ para $n \to \infty$, con la probabilidad 1; para los estados reversibles regresará un número ilimitado de veces a cada estado. El tiempo medio de regreso al estado inicial $i = 0$ es

$$\mu = \sum_{n=1}^{\infty} n v_n = (1 - p_0) + 2(1 - p_1) p_0 + 3(1 - p_2) p_0 p_1 + \ldots$$
$$\ldots + n(1 - p_{n-1}) p_0 \ldots p_{n-2} + \ldots =$$
$$= 1 + p_0 + p_0 p_1 + \ldots + p_0 p_1 \ldots p_{n-1} + \ldots,$$

y para la condición

$$\mu = 1 + \sum_{n=1}^{\infty} p_0 \ldots p_{n-1} < \infty$$

todos los estados son positivos.

En el caso más sencillo, cuando las probabilidades $1 - p_i$ de paso de todos los estados i al estado inicial 0 son tales que $1 - p_i \geqslant \delta > 0$, $i = 0, 1, \ldots$ se puede dar fácilmente la valuación del coeficiente de ergodicidad

$$k(1) > \inf_i (1 - p_i) \geqslant \delta$$

(véase (2.20)), y junto con ésta la valuación de la velocidad de convergencia hacia las probabilidades límites p_j^*, $j = 0, 1, \ldots$ (véase (2.22)), que forman la distribución estacionaria. Esta distribución satisface al sistema de ecuaciones (2.18) que en nuestro caso se repre-

senta del siguiente modo:

$$p_j^* = p_{j-1}^* p_{j-1}, \quad j = 1, 2, \ldots,$$

de donde

$$p_1^* = p_0^* p_0, \quad p_2^* = p_0^* p_0 p_1, \ldots, p_n^* = p_0^* p_0 \ldots p_{n-1}, \ldots$$

Ya que

$$1 = \sum_{n=0}^{\infty} p_n^* = p_0^* \Big[1 + \sum_{n=1}^{\infty} p_0 \ldots p_{n-1} \Big] = p_0^* \mu$$

entonces, como resultado tenemos

$$p_0^* = \frac{1}{\mu}, \quad p_1^* = \frac{p_0}{\mu}, \quad \ldots, \quad p_n^* = \frac{p_0 \ldots p_{n-1}}{\mu}, \quad \ldots \qquad (2.25)$$

§ 3. CADENAS DE MARKOV (TIEMPO CONTINUO)

1. Ecuaciones diferenciales para las probabilidades de paso. Examinemos el proceso homogéneo de Markov $\xi(t)$ con un número finito y numerable de estados 1, 2, ..., que se diferencia de las cadenas de Markov examinadas en el párrafo anterior, sólo, en que el parámetro t (tiempo) varía continuamente y el paso de un estado a otro es posible en cualquier momento de tiempo t. Igual que anteriormente, hablaremos convencionalmente sobre el sistema físico, cuyo estado fásico en el momento t es $\xi(t)$.

Designemos por $p_{ij}(t)$ la probabilidad de paso del estado i al estado j durante el tiempo t; precisemos que $p_{ij}(t)$ es, según la definición, la probabilidad de encontrarse en el estado j después del tiempo t, con la condición de que i fuese el estado inicial:

$$p_{ij}(t) = \mathbf{P}\{\xi(t+s) = j \mid \xi(s) = i\}, \quad i, j = 1, 2, \ldots$$

Sea p_i^0, $i = 1, 2, \ldots$, la distribución inicial de probabilidades. Análogamente a (2.1), para las probabilidades $p_j(t)$ de encontrarse en el estado correspondiente, $j = 1, 2, \ldots$, después del tiempo t tienen lugar las siguientes fórmulas:

$$p_j(t) = \sum_i p_i^0 p_{ij}(t), \quad j = 1, 2, \ldots, \qquad (3.1)$$

$$p_{ij}(t+s) = \sum_k p_{ik}(t)\, p_{kj}(s), \quad i, j = 1, 2, \ldots,$$

para todos los valores de t y s, donde

$$p_{ij}(0) = \begin{cases} 1 \text{ para } i = j \\ 0 \text{ para } i \neq j. \end{cases} \qquad (3.2)$$

Consideraremos que para todos los i, $j = 1, 2, \ldots$ las probabilidades de paso $p_{ij}(t)$ satisfacen las siguientes condiciones[1]) (compárese (2.5) y más adelante):

$$1 - p_{ii}(\Delta t) = \lambda_i \Delta t + o(\Delta t), \tag{3.3}$$
$$p_{ij}(\Delta t) = \lambda_{ij}\Delta t + o(\Delta t), \quad j \neq i,$$

donde $\frac{o(\Delta t)}{\Delta t} \to 0$ para $\Delta t \to 0$, λ_i es *la densidad de salida* del estado i; λ_{ij} es *la densidad de paso* de i al estado correspondiente j. Supongamos $\lambda_{ii} = -\lambda_i$, $i = 1, 2, \ldots$

Teorema 1. *Las probabilidades de paso p_{ij}, para el proceso de Markov con un número finito de estados para la condición* (3.3), *satisfacen a las ecuaciones diferenciales*

$$p'_{ij}(t) = \sum_k \lambda_{ik} p_{kj}(t), \quad i, j = 1, 2, \ldots, \tag{3.4}$$

y

$$p'_{ij}(t) = \sum_k p_{ik}(t)\lambda_{kj}, \quad i, j = 1, 2, \ldots, \tag{3.5}$$

con las condiciones iniciales (3.2).

Las ecuaciones (3.5) forman el llamado *sistema directo* y las ecuaciones (3.4), *el sistema inverso de ecuaciones diferenciales de Kolmogórov.*

Demostración. De acuerdo con la fórmula general (3.1),

$$p_{ij}(t + \Delta t) = \sum_k p_{ik}(\Delta t)\, p_{kj}(t) = \sum_k p_{ik}(t)\, p_{kj}(\Delta t).$$

Utilizando las expresiones asintóticas (3.3) obtenemos que

$$\frac{p_{ij}(t+\Delta t) - p_{ij}(t)}{\Delta t} = \sum_k \left[\lambda_{ik} + \frac{o(\Delta t)}{\Delta t}\right] p_{kj}(t) = \sum_k p_{ik}(t)\left[\lambda_{kj} + \frac{o(\Delta t)}{\Delta t}\right].$$

Las partes derechas de estas igualdades tienen un límite completamente determinado para $\Delta t \to 0$:

$$\lim_{\Delta t \to 0} \sum_k \left[\lambda_{ik} + \frac{o(\Delta t)}{\Delta t}\right] p_{kj}(t) = \sum_k \lambda_{ik} p_{kj}(t),$$
$$\lim_{\Delta t \to 0} \sum_k p_{ik}(t)\left[\lambda_{kj} + \frac{o(\Delta t)}{\Delta t}\right] = \sum_k p_{ik}(t)\lambda_{kj}.$$

Por consiguiente, también existe el límite

$$\lim_{\Delta t \to 0} \frac{p_{ij}(t+\Delta t) - p_{ij}(t)}{\Delta t} = p'_{ij}(t),$$

[1]) Señalemos que la relación (3.3) tendrá lugar, si (igual que en la deducción de la fórmula (2.6) para las probabilidades π_{ij} de paso desde i a j directamente) suponemos que la probabilidad de tal paso durante un tiempo pequeño Δt es $\lambda_{ij}\Delta t + o(\Delta t)$, y la probabilidad de más de un paso de estado a estado durante el tiempo Δt es $o(\Delta t)$.

donde deben cumplirse las igualdades (3.4) y (3.5). El teorema queda demostrado.

Se debe señalar que las ecuaciones diferenciales indicadas para las probabilidades de paso $p_{ij}(t)$ tienen lugar no sólo en el caso de un número finito de estados, sino también con algunas restricciones complementarias, y en el caso de número de estados numerables. Por ejemplo, el método anterior utilizado para la deducción del sistema directo de ecuaciones diferenciales queda en vigor, si en las expresiones asintóticas (3.3) los miembros restantes $o(\Delta t)$ son tales que

$$\frac{o(\Delta t)}{\Delta t} \to 0 \text{ para } \Delta t \to 0$$

es u n i f o r m e por todas las i, con ello, las densidades de paso λ_{ij} están limitadas u n i f o r m e m e n t e para el valor fijado de j:

$$\lambda_{ij} \leqslant \Lambda_j < \infty, \quad i = 1, 2, \ldots \tag{3.6}$$

Señalemos también que el sistema directo (3.5) tiene lugar en las probabilidades $p_j(t)$ para cualquier distribución inicial: precisamente de (3.1) y (3.5) se deduce inmediatamente que

$$p_j'(t) = \sum_k p_k(t)\,\lambda_{kj}, \quad j = 1, 2, \ldots \tag{3.7}$$

Examinemos algunos ejemplos.

Ejemplo (corriente de exigencias de Poisson).

Supongamos que en un sistema de servicio determinado en el transcurso del tiempo se reciben exigencias de modo que el número de exigencias $\xi(t)$ durante el tiempo t, es un proceso homogéneo de Markov (con los estados $i = 0, 1, \ldots$).

Desde el estado i sólo se puede pasar directamente al siguiente estado $i + 1$, $i = 0, 1, \ldots$. Supongamos que las densidades de salida desde cada estado i son las mismas e iguales a λ. Evidentemente, la probabilidad $p_{ii}(\Delta t)$ coincide, en nuestro caso, con la probabilidad de que el sistema se encuentre en el estado i en todo el intervalo de tiempo Δt, y de acuerdo con la fórmula general (1.7),

$$1 - p_{ii}(\Delta t) = \lambda \Delta t + o(\Delta t).$$

Está claro, que la densidad de salida λ coincide con la densidad de paso al siguiente estado, y de tal modo, en nuestro caso se cumplen las relaciones (3.3) en las cuales

$$\lambda_i = \lambda, \quad \lambda_{i,\,i+1} = \lambda, \quad \lambda_{ij} = 0 \text{ para } j \neq i,\ i + 1.$$

Para las probabilidades de paso $p_{ij}(t)$ tiene lugar el sistema de ecuaciones diferenciales (3.5). Evidentemente, $p_{ij}(t) = p_{0,\,j-i}(t)$. Supongamos que

$$p_j(t) = p_{0j}(t), \quad j = 0, 1, \ldots$$

Las ecuaciones diferenciales indicadas para las funciones $p_j(t)$ tienen la siguiente forma:

$$p_0'(t) = -\lambda p_0(t), \quad p_k'(t) = \lambda p_{k-1}(t) - \lambda p_k(t), \quad k = 1, 2, \ldots$$

Si pasamos a las nuevas funciones $f_k(t) = e^{\lambda t} p_k(t)$, entonces obtenemos que

$$f_0'(t) = \lambda f_0(t) + e^{\lambda t} p_0'(t) = \lambda f_0(t) - \lambda e^{\lambda t} p_0(t) = 0$$

$$f_k'(t) = \lambda f_k(t) + e^{\lambda t} p_k'(t) = \lambda f_k(t) + \lambda e^{\lambda t} p_{k-1}(t) - \\ - \lambda e^{\lambda t} p_k(t) = \lambda f_{k-1}(t) \quad k = 1, 2, \ldots,$$

donde $f_0(0) = 1$, $f_k(0) = 0$, para $k = 1, 2, \ldots$ El sistema de ecuaciones diferenciales de la forma

$$f_0'(t) = 0, \quad f_k'(t) = \lambda f_{k-1}(t), \quad k = 1, 2, \ldots,$$

con las condiciones iniciales indicadas tiene evidentemente, la siguiente solución:

$$f_0(s) = 1, \quad f_1(t) = \lambda t, \quad \ldots, \quad f_n(t) = \frac{(\lambda t)^n}{n!}, \quad \ldots$$

Volviéndonos a las funciones iniciales $p_k(t) = e^{-\lambda t} \cdot f_k(t)$ obtenemos que

$$p_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \ldots,$$

es decir, la corriente de exigencias examinada es una corriente *de Poisson*.

Ejemplo (sistema con tiempo exponencial de servicio). Supongamos que un sistema de servicio recibe una corriente de exigencias de Poisson con el parámetro λ_0 (las diferentes exigencias se reciben independientemente una de otra y la probabilidad de recibir una exigencia por separado durante el pequeño intervalo de tiempo Δt es $\lambda_0 \cdot \Delta t + o(\Delta t)$. Supongamos, que en el servicio de cada exigencia aislada se gasta un tiempo aleatorio τ, distribuido según una ley exponencial con el parámetro λ_1, es decir,

$$\mathbf{P}\{\tau > t\} = e^{-\lambda_1 t}$$

Examinemos dos estados del sistema de servicio; 0 significa el sistema libre, 1 significa el sistema ocupado. Consideraremos que si el sistema está ocupado, entonces las exigencias que se reciben en este tiempo, reciben la negativa y salen de la esfera de servicio.

Supongamos que en un momento t_0 el sistema se encuentra en el estado 0. Debido a la independencia del recibimiento de las exigencias, su comportamiento futuro no depende de las circunstancias anteriores, y en particular, durante el tiempo Δt el sistema pasa al estado 1 con la probabilidad $\lambda_0 \cdot \Delta t + o(\Delta t)$. Supongamos, que el sistema se

encuentra en el estado 1 en el momento t_1. Si designamos por τ_1 el momento aleatorio de paso de 1 a 0 (momento de finalizar el servicio), entonces, de acuerdo a la propiedad de la ley de distribución exponencial (véase el punto 2 del § 2, cap. II), para el tiempo de servicio tiene lugar la siguiente igualdad:

$$\mathbf{P}\{\tau_1 > t \mid \tau_1 > t_1\} = e^{-\lambda_1 (t-t_1)}.$$

Se ve que el paso al estado 0 y el comportamiento ulterior del sistema no dependen de su comportamiento hasta el momento t_1. En particular, durante el intervalo Δt siguiente a t_1, el sistema pasa al estado 0 con probabilidad $\lambda_1 \cdot \Delta t + o(\Delta t)$. De tal modo, la evolución del sistema se describe por el proceso de Markov con dos estados 0 y 1, y con las correspondientes densidades de paso λ_0, λ_1.

Examinemos la probabilidad de paso $p_{ij}(t)$. En nuestro caso $p_{01}(t) = 1 - p_{00}(t)$; $p_{10}(t) = 1 - p_{11}(t)$ y las ecuaciones diferenciales (3.5) tienen la forma

$$p'_{00}(t) + (\lambda_0 + \lambda_1)\, p_{00}(t) = \lambda_1,$$
$$p'_{11}(t) + (\lambda_0 + \lambda_1)\, p_{11}(t) = \lambda_0.$$

Resolviéndolas, obtenemos que

$$p_{00}(t) = \left(1 - \frac{\lambda_1}{\lambda_0 + \lambda_1}\right) e^{-(\lambda_0+\lambda_1)t} + \frac{\lambda_1}{\lambda_0 + \lambda_1},$$
$$p_{11}(t) = \left(1 - \frac{\lambda_0}{\lambda_0 + \lambda_1}\right) e^{-(\lambda_0+\lambda_1)t} + \frac{\lambda_0}{\lambda_0 + \lambda_1}.$$

Como conclusión de este punto señalemos que el sistema de ecuaciones diferenciales (3.5) puede tener solución diferente de las probabilidades de paso $p_{ij}(t)$ del proceso de Markov con las correspondientes densidades de paso λ_{ij}, si se altera *la condición de estabilidad* siguiente: *durante un tiempo finito ocurre con probabilidad 1 sólo un número finito de pasos.*

Ejemplo (proceso de multiplicación pura). Supongamos que el proceso homogéneo de Markov $\xi(t)$ describe el proceso de multiplicación de ciertas partículas ($\xi(t)$ es el número de partículas en el momento de tiempo t), en el cual, en caso de existir i partículas en el siguiente intervalo de tiempo Δt con probabilidad $\lambda_i \Delta t + o(\Delta t)$ se puede añadir una partícula, y la probabilidad de añadirle un número mayor de partículas es $o(\Delta t)$. Las densidades de paso λ_{ij} tienen la siguiente forma:

$$\lambda_{ij} = \begin{cases} \lambda_i \text{ para } j = i+1, \\ 0 \text{ para los restantes } j \end{cases} \qquad (i = 1, 2, \ldots).$$

Evidentemente, la condición de estabilidad consiste en que el tiempo τ, que se exige para un número infinito de pasos $1 \to 2 \to \ldots$ (desde 1 hasta ∞), debe ser infinito: $\tau = \infty$. Si introducimos las

magnitudes aleatorias τ_i, o sea, el tiempo de estancia en el estado i (τ_i coincide con el tiempo de espera del cambio de estado i), entonces tendremos

$$\tau = \tau_1 + \tau_2 + \ldots = \lim_{n\to\infty} \sum_{i=1}^{n} \tau_i,$$

donde τ_i, $i = 1, 2, \ldots$, son magnitudes aleatorias independientes, distribuidas según la ley exponencial con el parámetro correspondiente λ_i (véase (1.6)). La condición de estabilidad $\tau = \infty$ es equivalente a que

$$e^{-\tau} = \lim_{n\to\infty} \prod_{i=1}^{n} e^{-\tau_i} = 0.$$

Pero para la magnitud no negativa $e^{-\tau}$ la condición $e^{-\tau} = 0$ con probabilidad 1 es evidentemente equivalente, a que sea igual a cero la esperanza matemática $\mathbf{M}e^{-\tau}$, que es

$$\mathbf{M}e^{-\tau} = \lim_{n\to\infty} \mathbf{M}\left(\prod_{i=1}^{n} e^{-\tau_i}\right) = \lim_{n\to\infty} \prod_{i=1}^{n} \mathbf{M}e^{-\tau_i}$$

(véase el teorema 5 del § 4, cap. I). Tenemos

$$\mathbf{M}e^{-\tau_i} = \int_0^\infty e^{-t}(\lambda_i e^{-\lambda_i t})\,dt = \frac{\lambda_i}{1+\lambda_i} = 1 - \frac{1}{1+\lambda_i},$$

$$\mathbf{M}e^{-\tau} = \lim_{n\to\infty} \prod_{i=1}^{n} \left(1 - \frac{1}{1+\lambda_i}\right) = \prod_{i=1}^{\infty} \left(1 - \frac{1}{1+\lambda_i}\right),$$

y del tal modo, la condición de estabilidad examinada es equivalente a que

$$\prod_{i=1}^{\infty} \left(1 - \frac{1}{1+\lambda_i}\right) = 0.$$

Pero lo último significa que $\sum_{i=1}^{\infty} \frac{1}{1+\lambda_i} = \infty$ y esto a su vez es equivalente a la condición $\sum \lambda_i^{-1} = \infty$. De tal modo, la condición de estabilidad del proceso examinado de la multiplicación pura consiste en que

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty.$$

Supongamos que se altera esta condición; entonces, durante un intervalo de tiempo t, se forma un número infinito de partículas con una probabilidad positiva: $\mathbf{P}\{\tau < t\} < 0$. Supongamos que en el

momento inicial se tiene una sóla partícula. Introduzcamos un estado más, que significa la existencia de una cantidad infinita de partículas. El proceso pasa a este estado durante el tiempo t con la probabilidad positiva $p_{i\infty}(t) = \mathbf{P}\{\tau < t\}$, a pesar de que las densidades de paso correspondientes $\lambda_{i\infty}$ son todas iguales a cero y la ecuación diferencial (3.5) para $p_{i\infty}(t)$ tiene la forma

$$p'_{i\infty}(t) = 0, \quad p_{i\infty}(0) = 0;$$

su solución es la función $p_{i\infty}(t) \equiv 0$, mientras que la p r o b a b i l i d a d $p_{i\infty}(t) > 0$.

2. Coeficiente de ergodicidad y convergencia hacia la distribución estacionaria. Del mismo modo que el caso del tiempo discreto, tiene lugar el resultado siguiente.

Teorema 2. *Si el coeficiente de ergodicidad*

$$\boldsymbol{k}(t_0) = 1 - \frac{1}{2} \sup_{i,j} \sum_k |p_{ik}(t_0) - p_{jk}(t_0)|$$

es positivo para un determinado $t_0 > 0$, las probabilidades $p_j(t)$ de que el sistema se encuentre, después de un tiempo t, en el estado correspondiente $j = 1, 2, \ldots$, para $t \to \infty$, tienen los valores límites

$$p_j^* = \lim_{t\to\infty} p_j(t), \quad j = 1, 2, \ldots \tag{3.8}$$

Las probabilidades límites p_j^* no dependen de la distribución inicial p_i^0, $i = 1, 2, \ldots$, con esto

$$\sup_{p_i^0} |p_j(t) - p_j^*| \leqslant Ce^{-Dt} \tag{3.9}$$

$$\left(C = \frac{1}{1 - \boldsymbol{k}(t_0)}, \; D = \frac{1}{t_0} \ln \frac{1}{1 - \boldsymbol{k}(t_0)}\right).$$

(La demostración es la misma que en el teorema 4 del párrafo anterior).

Respecto a la condición $\boldsymbol{k}(t_0) > 0$ de este teorema, en el caso del tiempo continuo, se puede añadir lo siguiente: *si un estado j es accesible desde otro estado i (es decir, $p_{ij}(t_0) > 0$ para un determinado t_0), también es accesible para cualquier $t > 0$ con una probabilidad positiva $p_{ij}(t) > 0$* (para la cadena con un número finito de estados comunicantes, se deduce de aquí, que todas las probabilidades de paso y el coeficiente de ergodicidad $\boldsymbol{k}(t) \geqslant$ máx mín $p_{ij}(t)$ son rigurosamente positivos para todos los $t > 0$).

Demostremos esto para el caso de un número finito de estados. Examinemos al principio la probabilidad $p_{ii}(t)$ que se supone continua como función de t y $p_{ii}(0) = 1$, ya que esta probabilidad es positiva para valores de t suficientemente pequeños. Ahora bien, de la relación (3.1) se ve que para cualesquiera s y t

$$p_{ii}(s + t) \geqslant p_{ii}(s)\, p_{ii}(t),$$

y en realidad la probabilidad $p_{ii}(t)$ es positiva para todos los valores de t. Examinemos ahora el estado j accesible desde el estado i, es decir, tal estado que $p_{ij}(s) > 0$ para un determinado s. Mostremos que tal probabilidad $p_{ij}(t)$ es positiva para cualquier $t > 0$. Tenemos

$$p_{ij}(t) \geqslant p_{ij}(u)\, p_{jj}(t-u), \quad u \leqslant t.$$

Como hemos visto, la probabilidad $p_{jj}(t-u)$ es siempre positiva, de modo que es suficiente demostrar que $p_{ij}(u) > 0$ para $u \leqslant t$. Según la suposición, $p_{ij}(s) > 0$. Examinemos la cadena de Markov con el tiempo discreto y con las probabilidades de paso $p_{ij} = p_{ij}\left(\frac{s}{n}\right)$, donde el número natural n satisface a la condición $m\frac{s}{n} \leqslant t$ (m es el número de estados). Ya que $p_{ij}\left(n\frac{s}{n}\right) > 0$, el estado j es accesible desde i. Se ve fácilmente, que es accesible no sólo después de n pasos, sino también en un determinado número de pasos n_0, no superior al número m de todos los estados existentes, es decir, $p_{ij}\left(n_0\frac{s}{n}\right) > 0$, donde $n_0\frac{s}{n} = u \leqslant t$. La demostración está terminada.

Es útil señalar la siguiente circunstancia. Eligiendo una unidad convencional de tiempo Δ, se puede considerar a $\xi(t)$, en el momento $t = k\Delta$, como una cadena de Markov $\tilde{\xi}(k)$, $k = 0, 1, \ldots$, con las probabilidades de paso $\tilde{p}_{ij}(k) = p_{ij}(k\Delta)$. Evidentemente, las probabilidades límites p_j^*, $j = 1, 2, \ldots$, serán las mismas que para el proceso inicial $\xi(t)$. Si referimos los estados $j = 1, 2, \ldots$, de este proceso al mismo tipo que en la cadena correspondiente de Markov $\tilde{\xi}(k)$ (cada estado j puede ser reversible o irreversible, positivo o nulo), entonces de la relación (3.8), en el caso de tiempo continuo t (lo mismo que de la relación (2.21) en el caso del tiempo discreto t) se deduce, que las probabilidades límites (3.8), *para la clase cerrada de estados positivos, forman la única distribución estacionaria*

$$p_j^* = \sum_i p_i^* p_{ij}(t), \quad j = 1, 2, \ldots, \tag{3.10}$$

para todos los valores de t.

Para las probabilidades estacionarias $p_j(t) = p_j^*$, obtenemos de las ecuaciones diferenciales (3.7) (para $p_j'(t) = 0$) el siguiente sistema lineal:

$$\sum_i p_i^* \lambda_{ij} = 0, \quad j = 1, 2, \ldots \tag{3.11}$$

Ejemplo (*sistema de servicio. Fórmula de Erlang*). Examinemos el sistema que puede atender a la vez a m demandas. Consideraremos que tenemos m líneas y la demanda siguiente se recibe en una de las líneas, si por lo menos está libre una de ellas; en caso contrario a la demanda

recibida se le da la negativa y aquélla sale fuera de la esfera de servicio. Supongamos, que la corriente de demandas es la corriente de Poisson con el parámetro λ_0, y que las demandas son atendidas independientemente y que el tiempo de servicio en cada demanda (en cada una de las m líneas) está distribuido según la ley exponencial con el parámetro λ.

Examinemos los estados $k = 0, 1, \ldots, m$, donde el estado k significa que están ocupadas justamente k líneas. El paso del sistema de estado a estado durante el tiempo t, representa en sí un proceso de Markov, cuya densidad de paso tiene la forma

$$\lambda_{0j} = \begin{cases} \lambda_0 \text{ para } j = 1, \\ 0 \text{ para } j \neq 1 \end{cases} \qquad \lambda_{kj} = \begin{cases} \lambda_0 \text{ para } j = k+1, \\ k\lambda \text{ para } j = k-1, \\ 0 \text{ para } j \neq k-1,\ k+1. \end{cases}$$

En efecto, el paso desde k a $k+1$ se realiza al recibirse la demanda siguiente, que ocurre en el tiempo Δt con la probabilidad $\lambda_0 \Delta t + o(\Delta t)$. La probabilidad de que ninguna de las k líneas ocupadas se libere en el tiempo Δt es $[1 - \lambda \Delta t - o(\Delta t)]^k$ (ya que las líneas se sirven, independientemente una de otra) y la probabilidad de liberación de una de las líneas, es decir, el paso del estado k al $k-1$ es $1 - [1 - \lambda \Delta t - o(\Delta t)]^k = k\lambda \Delta t + o(\Delta t)$. La probabilidad de otros cambios en el sistema durante el intervalo de tiempo Δt es $o(\Delta t)$.

Como se ve directamente de la expresión para las probabilidades $p_{ij}(t)$, obtenidas antes en el caso $m = 1$ (véase la pág. 165), cuando $t \to \infty$ las probabilidades $p_{ij}(t)$ tienden rápido, exponencialmente, hacia sus valores límites. De acuerdo a la valuación general (3.9), también se observa lo mismo en el caso de m líneas. Las probabilidades estacionarias $p_j^*(t)$ pueden ser halladas de las ecuaciones (3.11), que en nuestro caso tienen la siguiente forma:

$$-\lambda_0 p_0^* + \lambda p_1^* = 0,$$

$$\lambda_0 p_{k-1}^* - (\lambda_0 + k\lambda)\, p_k^* + (k+1)\, \lambda p_{k+1}^* = 0, \quad 1 \leqslant k < m,$$

$$\lambda_0 p_{m-1}^* - m\lambda p_m^* = 0.$$

De estas ecuaciones obtenemos que

$$p_k = \frac{\frac{1}{k!}\left(\frac{\lambda_0}{\lambda}\right)^k}{\sum_{l=0}^{m} \frac{1}{l!}\left(\frac{\lambda_0}{\lambda}\right)^l}, \quad k = 0, 1, \ldots, m.$$

(Las expresiones halladas para las probabilidades estacionarias se denominan *fórmulas de Erlang*).

§ 4. PROCESOS QUE SE RAMIFICAN

1. Ecuación diferencial de las funciones generatrices. Supongamos que se tiene un conjunto determinado de partículas, que en el transcurso del tiempo se transforman en partículas del mismo tipo, con ello este proceso de «multiplicación» posee la siguiente propiedad: cada una de las partículas iniciales, después de un tiempo t, independientemente de las otras partículas y de las circunstancias precedentes al momento inicial, genera un grupo de n partículas con la misma probabilidad $p_n(t)$ para todas las partículas.

Designemos por $\xi(t)$ el número de partículas existentes en el momento de tiempo t. Evidentemente, la evolución de la magnitud $\xi(t)$ representa en sí un proceso aleatorio de Markov. Un proceso de tal tipo se llama *proceso que se ramifica*[1]).

Supongamos que en un momento de tiempo s, digamos $s = 0$, se tienen justamente k partículas. Designemos por $\xi_i(t)$ el número de partículas generadas por la partícula i, $i = 1, \ldots, k$, después de un tiempo t. Entonces el número total de partículas después de un tiempo t será

$$\xi(t) = \xi_1(t) + \ldots + \xi_k(t). \tag{4.1}$$

Aquí, las magnitudes aleatorias $\xi_1(t), \ldots, \xi_k(t)$ son independientes entre sí y tienen una misma distribución de probabilidades

$$\mathbf{P}\{\xi_i(t) = n\} = p_n(t), \quad n = 0, 1, \ldots$$

Supongamos que una partícula aislada, en el pequeño intervalo de tiempo Δt con la probabilidad

$$p_n(\Delta t) = \lambda_n \cdot \Delta t + o(\Delta t), \quad n \neq 1,$$

se transforma en n partículas nuevas, y se queda invariable, con la probabilidad

$$p_1(\Delta t) = 1 - \lambda \Delta t + o(\Delta t).$$

Supongamos luego, que $\lambda_1 = -\lambda$, $\sum_k \lambda_k = 0$ y las probabilidades de paso $p_n(t) = p_{1n}(t)$ satisfacen a la ecuación diferencial de Kolmogórov (véase (3.4))

$$\frac{d}{dt} p_n(t) = \sum_k \lambda_k p_{kn}(t), \quad n = 0, 1, \ldots,$$

[1]) El modelo descrito puede servir para estudiar muchos procesos reales (reacciones fotoquímicas, procesos nucleares, etc.).

donde $p_{kn}(t)$ son las probabilidades de paso del proceso que se ramifica de Markov $\xi(t)$ ($p_{kn}(t)$ es la probabilidad de que, k partículas generen n partículas durante el tiempo t).

Introduzcamos las funciones generatrices

$$F(t, z) = \sum_{n=0}^{\infty} p_n(t) z^n,$$
$$F_k(t, z) = \sum_{n=0}^{\infty} p_{kn}(t) z^n. \tag{4.2}$$

Para cada z, $|z| < 1$, tenemos

$$\sum_{n=0}^{\infty} z^n \frac{d}{dt} p_n(t) = \frac{d}{dt} \sum_{n=0}^{\infty} p_n(t) z^n = \sum_k \lambda_k \sum_{n=0}^{\infty} p_{kn}(t) z^n,$$

lo que da la siguiente ecuación diferencial para la función generatriz $F(t, z)$:

$$\frac{d}{dt} F(t, z) = \sum_k \lambda_k F_k(t, z). \tag{4.3}$$

Las funciones $F(t, z)$ y $F_k(t, z)$ determinadas por las fórmulas (4.2) son tales que para un valor fijado de z representan en sí las esperanzas matemáticas

$$F(t, z) = \mathbf{M} z^{\xi_i(t)}, \quad i = 1, \ldots, k,$$
$$F_k(t, z) = \mathbf{M} z^{\xi(t)},$$

donde $\xi_i(t)$ es el número de partículas generadas por la partícula i durante el tiempo t y $\xi(t) = \sum_{i=1}^{k} \xi_i(t)$. Ya que todas las magnitudes $\xi_i(t)$, $i = 1, \ldots, k$ son independientes, entonces

$$\mathbf{M} z^{\xi_1(t) + \ldots + \xi_k(t)} = \mathbf{M} z^{\xi_1(t)} \ldots \mathbf{M} z^{\xi_k(t)},$$

lo que da la siguiente relación para las funciones generatrices examinadas

$$F_k(t, z) = [F(t, z)]^k, \quad k = 1, 2, \ldots \tag{4.4}$$

Ya que $F_0(t, z) \equiv 1$, la ecuación diferencial para la función productora $F(t, z)$ se puede escribir en la forma

$$\frac{d}{dt} F(t, z) = \sum_k \lambda_k F^k(t, z). \tag{4.5}$$

Consideraremos que las densidades de paso λ_k, $k = 0, 1, \ldots$ son parámetros dados del proceso que se ramifica $\xi(t)$. Introduzcamos

la función

$$f(x) = \sum_{k=0}^{\infty} \lambda_k x^k, \tag{4.6}$$

la cual es analítica para $0 < x < 1$ y la fórmula (4.6) da su desarrollo en serie de potencias. De acuerdo a la igualdad (4.5) la función generatriz $F(t, z)$ es la solución de la ecuación diferencial de la forma

$$\frac{dx}{dt} = f(x). \tag{4.7}$$

Ya que $F(0, z) = z$, la función generatriz $F(t, z)$ para cada z, $0 \leqslant z \leqslant 1$, coincide con la solución $x = x(t)$ de esta ecuación que satisface a la condición inicial $x(0) = z$.

En lugar de la ecuación (4.7), es cómodo examinar su ecuación diferencial equivalente para la función $t = t(x)$ inversa a $x = x(t)$

$$\frac{dt}{dx} = \frac{1}{f(x)}.$$

La función $t = t(x)$, que es la solución de esta ecuación, tiene la forma

$$t = \int_z^x \frac{du}{f(u)}, \quad 0 \leqslant x \leqslant 1. \tag{4.8}$$

Ejemplo. Supongamos que las densidades de paso son

$$\lambda_0 = \lambda, \quad \lambda_1 = -\lambda, \quad \lambda_k = 0 \quad \text{para} \quad k = 2, 3, \ldots$$

En este caso $f(x) = \lambda(1 - x)$ y

$$t = \int_z^x \frac{du}{f(u)} = -\frac{1}{\lambda}[\ln(1-x) - \ln(1-z)].$$

De esta relación se obtiene fácilmente la función $F = F(t, z)$. Precisamente,

$$\ln(1 - F) = -\lambda t + \ln(1 - z)$$

y

$$F(t, z) = 1 - e^{-\lambda t}(1 - z).$$

Las probabilidades $p_n(t)$ determinadas del desarrollo $F(t, z) = \sum_{n=0}^{\infty} p_n(t) z^n$ en el caso examinado son

$$p_0(t) = 1 - e^{-\lambda t}, \quad p_1(t) = e^{-\lambda t}, \quad p_n(t) = 0 \quad \text{para} \quad n = 2, 3, \ldots$$

Ejemplo. Imaginémonos que algunas partículas se multiplican mediante su fisión por la mitad; en el proceso que se ramifica, co-

rrespondiente a este fenómeno, hay que tomar

$$\lambda_0 = 0, \quad \lambda_1 = -\lambda, \quad \lambda_2 = \lambda \quad (\lambda_k = 0 \text{ para } k < 2).$$

En este caso

$$f(x) = \lambda x(x-1)$$

y

$$t = \int_z^x \frac{du}{f(u)} = \frac{1}{\lambda} \int_z^x \left[\frac{1}{u-1} - \frac{1}{u}\right] du = \frac{1}{\lambda}\left[\ln\left(1-\frac{1}{x}\right) - \ln\left(1-\frac{1}{z}\right)\right].$$

De tal modo, la función $F(t, z) = \sum_{n=0}^{\infty} p_n(t) z^n$ puede ser determinada de la relación

$$t = \frac{1}{\lambda}\left[\ln\left(1-\frac{1}{F}\right) - \ln\left(1-\frac{1}{z}\right)\right].$$

Hallamos fácilmente que

$$F(t, z) = e^{-\lambda t} \frac{z}{1 - z(1-e^{-\lambda t})} = e^{-\lambda t} z \sum_{k=0}^{\infty} (1-e^{-\lambda t})^k z^k.$$

Como resultado obtenemos que

$$p_0(t) = 0, \quad p_n(t) = e^{-\lambda t}(1-e^{-\lambda t})^{n-1}, \quad n = 1, 2, \ldots$$

Examinemos la ecuación diferencial (4.7) en la cual la función $f(x)$ se determina por la fórmula (4.6). De esta fórmula se ve que

$$f''(x) = \sum_{k=2}^{\infty} k(k-1)\lambda_k x^{k-2} \geqslant 0$$

para $0 \leqslant x < 1$,

de modo que la función $f(x)$ es convexa y su derivada $f'(x)$ crece monótonamente en el intervalo $0 \leqslant x \leqslant 1$. El valor de $x = 1$ es una raíz de la ecuación $f(x) = 0$, ya que $\sum_k \lambda_k = 0$. Es posible que sólo exista todavía otra raíz $x = \alpha$ de esta ecuación, y en correspondencia con esto, el gráfico de la función $f(x)$ tiene la forma indicada en la fig. 20.

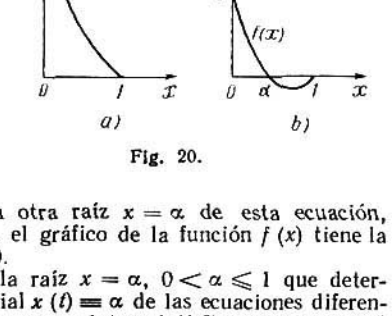

Fig. 20.

Supongamos que se tiene la raíz $x = \alpha$, $0 < \alpha \leqslant 1$ que determina una curva integral especial $x(t) \equiv \alpha$ de las ecuaciones diferenciales examinadas. Tomemos la curva integral (4.8) que pasa por el

punto $t = 0$, $x = z$, $0 \leqslant z < \alpha$:

$$t = \int_z^x \frac{du}{f(u)}.$$

Ya que la derivada $f'(\alpha)$ es finita y para $x \sim \alpha$ la función $f(x)$ tiene la forma $f(x) \sim f'(\alpha) \cdot (x - \alpha)$, entonces a lo largo de la curva integral crece ilimitadamente el valor $t = \int_z^x \frac{dx}{f(u)}$ para $x \to \alpha$, con ello la propia curva no corta en ningún lugar a otra curva integral $x(t) \equiv \alpha$. En el intervalo $0 \leqslant x < \alpha$ la función $f(x)$ es positiva y, por consiguiente, la curva integral $x = x(t)$ es positiva y, por consiguiente, la curva integral $x = x(t)$ crece monótonamente para $t \to \infty$, quedando limitada por el valor $x = \alpha$. Como función monótona limitada, la función $x(t)$ tiene un límite determinado $\beta = \lim_{t \to \infty} x(t)$, $z \leqslant \beta \leqslant \alpha$. Pero cuando $x \to \beta$ la función $f(x)$ tiene su límite $f(\beta)$:

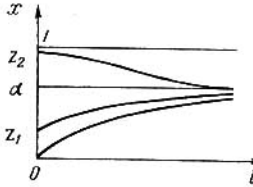

Fig. 21.

$$f(\beta) = \lim_{t \to \infty} f[x(t)] = \lim_{t \to \infty} x'(t).$$

Está claro que el valor $f(\beta)$ deberá ser igual a cero, ya que en caso contrario la función

$$x(t) = z + \int_0^t f[x(s)]\, ds$$

crecerá ilimitadamente para $t \to \infty$. Por consiguiente, β es raíz de la ecuación $f(x) = 0$, y coincide con α: $\beta = \alpha$. De tal modo, todas las curvas integrales $x = x(t)$ para $t = 0$ que pasan por el punto $x = z$, $0 \leqslant z < \alpha$ crecen monótonamente para $t \to \infty$ y

$$\lim_{t \to \infty} x(t) = \alpha. \tag{4.9}$$

Es completamente análogo el comportamiento de las curvas integrales que pasan por los puntos $x = z$, $\alpha < z < 1$ $(0 \leqslant \alpha < 1)$ para $t = 0$. La diferencia sólo consiste en que $x(t)$ decrece monótonamente, ya que la derivada $x'(t) = f[x(t)]$ es negativa $(f(x) \leqslant 0$ para $\alpha < x < 1)$. El cuadro general de las curvas integrales correspondientes a los valores del parámetro z en el intervalo $0 \leqslant z < 1$, está expuesta en la fig. 21.

El caso $z = 1$ exige un examen especial. A él siempre le corresponde la curva integral de la forma $x(t) \equiv 1$.

Si para un determinado x_0, $\alpha < x_0 < 1$,

$$\int_{x_0}^{1} \frac{dx}{f(x)} = -\infty \tag{4.10}$$

(que siempre es así, cuando $f'(1) < \infty$), entonces la curva integral arbitraria de la forma

$$t = t_0 + \int_{x_0}^{x} \frac{du}{f(u)}, \quad 0 \leqslant x < 1,$$

que pasa por un punto determinado (t_0, x_0), para $x \to 1$, decrece ilimitadamente:

$$t = t_0 + \int_{x_0}^{x} \frac{du}{f(u)} \to -\infty.$$

Esto quiere decir, que cualquiera que fuese $t_0 > 0$, para un determinado $x = z$, $0 \leqslant z < 1$ tiene lugar la igualdad

$$t(z) = t_0 + \int_{x_0}^{z} \frac{du}{f(u)} = 0.$$

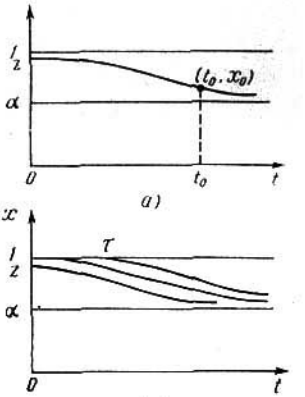

Fig. 22.

Vemos (véase la fig. 22,*a*) que todas las curvas integrales cortan al eje $t = 0$ en un punto $(0, z)$, donde $0 \leqslant z < 1$ y, por consiguiente, $x(t) \equiv 1$ es la única curva integral que pasa por el punto $(0, 1)$.

Si

$$\int_{x_0}^{1} \frac{dx}{f(x)} > -\infty, \tag{4.11}$$

entonces para valores suficientemente grandes de $t_0 > 0$, la curva integral $t = t_0 + \int_{x_0}^{x} \frac{dx}{f(x)}$ corta a la curva integral $x(t) \equiv 1$ siendo tangente a ésta en un punto determinado $(\tau, 1)$, donde

$$\tau = t_0 + \int_{x_0}^{1} \frac{dx}{f(x)}$$

(véase la fig. 23, *b*). En este caso, por el punto $(0, 1)$ pasa toda una familia de curvas integrales $x_\tau(t)$, cada una de las cuales corresponde

a su valor $\tau \geqslant 0$. Entre ellas se encuentra la curva integral $x_0(t)$ correspondiente al valor $\tau = 0$ y teniendo las mismas propiedades que la curva $x_0(t)$ se halla por debajo de todas las curvas integrales restantes $x_\tau(t)$:

$$x_0(t) \leqslant x_\tau(t), \quad 0 \leqslant t < \infty.$$

Esto se explica, porque dentro de la zona $0 \leqslant x < 1$, $0 < t < \infty$, la solución de la ecuación diferencial examinada es única y en esta zona las curvas integrales no se cortan entre sí. También se ve fácilmente que la curva integral $x_0(t)$ sirve de límite para las otras curvas integrales $x(t, z)$ que descansan debajo de ella y pasan por el punto correspondiente $(0, z)$, donde $0 \leqslant z < 1$

$$x_0(t) = \lim_{z \to 1} x(t, z). \tag{4.12}$$

2. Efectos de degeneración y explosión. Recordemos que la función generatriz $F(t, z) = \sum_{k=0}^{\infty} p_k(t) z^k$ del proceso que se ramifica (donde $p_k(t) = \mathbf{P}\{\xi(t) = k\}$ a condición de que $\xi(0) = 1$), como función de t, coincide con la solución $x(t, z)$ de la ecuación diferencial (4.7), con la condición inicial $x(0, z) = z$. El análisis de esta ecuación diferencial expuesto anteriormente, permite hacer las siguientes conclusiones referentes al propio proceso que se ramifica $\xi(t)$.

Hablando en general, existe la probabilidad positiva de que después de un tiempo determinado t no quede ni una sola partícula. Naturalmente, esto no puede ocurrir, si $\lambda_0 = 0$ (es decir, si las partículas no pueden desaparecer sino solamente multiplicarse). Si en el momento inicial $t = 0$ se tiene una partícula, entonces esta probabilidad es $p_0(t) = F(t, 0)$. Si al principio se tienen k partículas, esta probabilidad será $p_{k0}(t) = F^k(t, 0) = p_0^k(t)$.

La probabilidad $p_0(t)$, como función de t, es una solución de la ecuación diferencial (4.7) correspondiente al parámetro $z = 0$:

$$p_0'(t) = f[p(t)], \quad p_0(0) = 0.$$

Anteriormente fue mostrado que esta solución, cuando $t \to \infty$, se aproxima asintóticamente a un valor determinado $p_0 = \alpha$ que es la raíz menor de la ecuación $f(x) = 0$ (véase la relación (4.9)), es decir,

$$\lim_{t \to \infty} p_0(t) = \alpha. \tag{4.13}$$

De tal modo, $p_0 = \alpha$ *es la probabilidad de degeneración del proceso que se ramifica* $\xi(t)$ (la probabilidad de que en un cierto momento de tiempo no quede ni una sola partícula).

Si la función $f(x)$ es positiva en el intervalo $0 \leqslant x < 1$, la probabilidad de degeneración del proceso que se ramifica $\xi(t)$ será igual a 1.

Examinemos el llamado *fenómeno de explosión*, cuando se forma una cantidad infinita de partículas.

La probabilidad de que ocurra la explosión antes del momento t (si al principio había una partícula) es

$$p_\infty(t) = 1 - \mathbf{P}\{\xi(t) < \infty\} =$$

$$= 1 - \sum_{n=0}^{\infty} \mathbf{P}(\xi(t) = n) = 1 - \sum_{n=0}^{\infty} p_n(t) = 1 - \lim_{z \to 1} F(t, z).$$

En el caso cuando $x(t) \equiv 1$ es la única curva integral de la ecuación diferencial (4.7) que pasa por el punto (0, 1), el límite $\lim_{z \to 1} F(t, z)$ evidentemente, será igual a 1. Por consiguiente, con la condición (4.10), $p_\infty(t) = 0$ para cualquier t, de modo que la probabilidad de la explosión queda excluida. Para la condición (4.11), las curvas integrales $x(t, z)$ cuando $z \to 1$, convergen monótonamente hacia la función $x_0(t)$ descrita anteriormente (véase la relación (4.12)) de modo que *la probabilidad de explosión* es

$$p_\infty(t) = 1 - x_0(t) > 0; \tag{4.14}$$

Se ve fácilmente, que si al principio había k partículas, entonces la probabilidad de explosión después del tiempo t para la condición (4.10) también será igual a cero, y para la condición (4.11) será

$$p_{k\infty}(t) = 1 - x_0^k(t). \tag{4.15}$$

§ 5. ALGUNOS PROCESOS DE SERVICIOS EN MASA Y FLUCTUACIONES ALEATORIAS

1. Procesos de restablecimiento. Sea $\xi_1, \xi_2, \ldots$ una sucesión de magnitudes aleatorias positivas independientes, que tienen la misma distribución de probabilidades, y

$$S_n = \xi_1 + \xi_2 + \ldots + \xi_n.$$

Consideraremos convencionalmente, que se tiene un cierto aparato con el plazo de servicio ξ_1; después de salir fuera de uso (después del tiempo aleatorio ξ_1) se cambia por otro aparato nuevo, que a su vez sale fuera de uso después del tiempo aleatorio ξ_2, después de lo cual se cambia por el siguiente aparato nuevo, etc. Con tal interpretación, las magnitudes S_n, $n = 1, 2, \ldots$, se denominan naturalmente *momentos de restablecimiento*.

En *el proceso de restablecimiento* examinado S_n, $n = 1, 2, \ldots$, nos ocuparemos, fundamentalmente, de las magnitudes $\nu(t)$ que es

el número de restablecimientos en el intervalo de tiempo $[0, t]$ y $\eta(t)$, que es el plazo restante de servicio del aparato sucesivo que trabaja en el momento t.

Supondremos que el restablecimiento se realiza en momentos discretos de tiempo $t = kh$ múltiplos de un cierto $h > 0$ (digamos $h = 1$). En relación con esto debemos precisar en que momento se reemplaza el aparato estropeado. Examinando sólo los valores de números enteros $t \geqslant 0$, consideraremos que si el aparato se estropeó en el intervalo $[t, t+1]$, se cambia por otro aparato nuevo en el momento de tiempo $t+1$ (el plazo de servicio de cada aparato puede ser igual a 1, 2, . . .).

Así pues, supongamos que las magnitudes examinadas ξ_k, $k = 1, 2, \ldots$, *toman los valores de números enteros* $x = 1, 2, \ldots$ con las probabilidades correspondientes $P(x)$, $\sum_1^\infty P(x) = 1$.

Designemos por $\xi(t)$ el tiempo de trabajo del aparato *siguiente* en el momento t:

$$\xi(t) = t - S_{\nu(t)}. \tag{5.1}$$

En el siguiente momento de tiempo $t+1$ el aparato bien sale fuera de trabajo y se cambia por un aparato nuevo, en este caso $\xi(t+1) = 0$, o bien continúa trabajando, en este caso $\xi(t+1) = \xi(t) + 1$. Con la condición de que el aparato que trabajó el tiempo x, sale fuera de trabajo después de la unidad de tiempo y se cambia por un aparato nuevo con la probabilidad condicional

$$q(x) = \mathbf{P}\{\xi_1 = x+1 \mid \xi_1 > x\} = \frac{P(x+1)}{\sum_{y=x+1}^{\infty} P(y)}, \quad x = 0, 1, \ldots \tag{5.2}$$

El proceso aleatorio

$$\xi(0) \to \xi(1) \to \xi(2) \ldots$$

representa en sí la cadena de Markov del tipo examinado anteriormente (véase, el ejemplo de la pág. 159) cuando es posible el paso del estado x, al estado siguiente $x+1$ que se realiza con la probabilidad

$$p(x) = 1 - q(x) = \frac{G(x+1)}{G(x)}, \quad x = 0, 1, \ldots, \tag{5.3}$$

o bien al estado «inicial» 0, que se realiza con la probabilidad $q(x)$ determinada por la fórmula (5.2); en la expresión (5.3) $G(x) = \sum_{y=x+1}^{\infty} P(y)$ significa la probabilidad de que el plazo de servicio del aparato sea mayor que x Evidentemente, la caída en el estado 0

durante t ($\xi(t) = 0$) significa el restablecimiento en el momento t; el tiempo de regreso al estado 0 no es otra cosa que el tiempo de servicio de un aparato; el número de caídas en el estado 0 durante el intervalo $(s, t]$ es igual al número de restablecimientos en el intervalo de tiempo desde s hasta t, etc.

Como fue establecido para nuestra cadena de Markov $\xi(t)$ con los estados $x = 0, 1, \ldots$, con la condición de que el tiempo medio de regreso μ al estado 0 sea finito se tiene *la distribución estacionaria de probabilidades* (2.25):

$$p^*(0) = \frac{1}{\mu}\ ;\ p^*(x) = \frac{p(0)\,p(1)\ldots p(x-4)}{\mu}\ ,\quad x = 1, 2, \ldots,$$

que para las probabilidades de paso $p(x) = \frac{G(x+1)}{G(x)}$, $x = 0, 1, \ldots$, tiene la forma

$$p^*(x) = \frac{G(x)}{\mu}\ ,\quad x = 0, 1, \ldots\ ; \tag{5.4}$$

con ello *el tiempo medio de regreso* $\mu = \sum_{x=0}^{\infty} G(x)$ *coincide con el tiempo medio de servicio del aparato aislado*:

$$\sum_{x=1}^{\infty} xP(x) = \sum_{x=1}^{\infty} x[G(x-1) - G(x)] =$$

$$= [G(0) - G(1)] + 2[G(1) - G(2)] + 3[G(2) - G(3)] + \ldots =$$

$$= \sum_{x=0}^{\infty} G(x) = \mu. \tag{5.5}$$

Recordemos que $G(x)$ es la probabilidad de que el plazo de servicio del aparato sea mayor que x; supongamos que $G(x)$ decrece cuando $x \to \infty$ lo suficientemente rápido, a saber:

$$G(x+1) \leqslant \alpha G(x), \quad \text{donde} \quad \alpha < 1. \tag{5.6}$$

En este caso todas las probabilidades de paso $q(x) = 1 - \frac{G(x+1)}{G(x)}$ no son menores que $1 - \alpha > 0$, y por consiguiente, *el coeficiente de ergodicidad* de tal cadena de Markov será positivo:

$$k(1) \geqslant 1 - \alpha.$$

Con esta condición las probabilidades $p_t(x) = \mathbf{P}\{\xi(t) = x\}$ para $t \to \infty$, convergen hacia la probabilidad estacionaria $p^*(x)$:

$$\lim_{t\to\infty} p_t(x) \frac{G(x)}{\mu}\ ,\quad x = 0, 1, \ldots, \tag{5.7}$$

y aún más, tiene lugar la siguiente valuación de la velocidad de convergencia (véase el teorema 4 del § 2 de este capítulo)

$$\left| p_t(x) - \frac{G(x)}{\mu} \right| \leqslant \frac{1}{\alpha} e^{-\alpha t}. \tag{5.8}$$

Examinemos el número medio de restablecimientos durante el tiempo desde t hasta $t + s$ (de otro modo, el número medio de caídas en el estado 0 durante ese tiempo), que es

$$n(t+s) - n(t) = \mathbf{M}[\nu(t+s) - \nu(t)] = \sum_{u=t+1}^{t+s} p_u(0)$$

(anteriormente designamos por $\nu(t)$ al número de restablecimientos en el intervalo de tiempo de 0 a t). Está claro, que para cualquier intervalo limitado $(t, t+s]$

$$n(t+s) - n(t) \to \frac{s}{\mu} \quad \text{para} \quad t \to \infty, \tag{5.9}$$

ya que $p_u(0) \to \frac{1}{\mu}$ y, además, como se deduce de la desigualdad (5.8),

$$\left| [n(t+s) - n(t)] - \frac{s}{\mu} \right| \leqslant \frac{s}{\alpha} e^{-\alpha t}. \tag{5.10}$$

La relación (5.9) muestra que *para un proceso prolongado de restablecimiento el número medio de restablecimientos durante el tiempo s es aproximadamente igual a la magnitud* $\frac{s}{\mu}$, *donde* μ *es el tiempo medio de servicio del aparato aislado.*

Hemos convenido en considerar que si el aparato trabajó en el momento t (múltiplo de h) entonces él se cambia no antes de un tiempo h ($h = 1$). Como tiempo durante el cual todavía trabaja el aparato después del momento t, tomamos la magnitud

$$\eta(t) = S_{\nu(t)+1} - h - t, \tag{5.11}$$

donde $S_{\nu(t)+1}$ es el momento de restablecimiento siguiente a t (momento de cambio del aparato examinado).

Está intuitivamente claro, que durante el proceso prolongado de restablecimiento ($t \to \infty$) la distribución de probabilidades de la magnitud $\eta(t)$ deberá ser aproximadamente igual a la de la magnitud $\xi(t) = t - S_{\nu(t)}$. De base para ello sirve la «reversibilidad» durante el tiempo del proceso de restablecimiento; los momentos de restablecimiento se pueden contar en orden contrario (el número de restablecimientos $n = \nu(t)$ crece ilimitadamente, y las magnitudes

$$S_1 = \xi_1, \; S_2 = \xi_1 + \xi_2, \; \ldots, \; S_n = \xi_1 + \ldots + \xi_n$$

y

$$S'_1 = \xi_n, \; S'_2 = \xi_n + \xi_{n+1}, \; \ldots, \; S'_n = \xi_n + \ldots + \xi_1$$

tienen la misma distribución conjunta).

Señalemos, que en el caso cuando el tiempo de espera de la rotura del aparato tiene la distribución exponencial de probabilidades

$$G(x) = e^{-\lambda x}, \quad x = 0, 1, \ldots,$$

ya desde el principio

$$\mathbf{P}\{\eta(t) = x\} = \frac{G(x)}{\mu}, \quad x = 0, 1, \ldots \left(\mu = \frac{1}{1 - e^{-\lambda}}\right)$$

(véase el punto 2 del § 2, cap. II).

Sin limitarnos a las observaciones hechas anteriormente, damos un fundamento riguroso al hecho de que *la magnitud* $\eta(t)$ *para* $t \to \infty$, *tiene la distribución límite de probabilidades*

$$\lim_{t\to\infty} \mathbf{P}\{\eta(t) = x\} = \frac{G(x)}{\mu}, \quad x = 0, 1, \ldots \tag{5.12}$$

Suponiendo que $S_0 = 0$, tenemos

$$\mathbf{P}\{\eta(t) = x\} = \sum_{n=0}^{\infty} \mathbf{P}\{S_n \leqslant t,\ S_{n+1} = t + 1 + x\},$$

donde (según la fórmula de la probabilidad total)

$$\mathbf{P}\{S_n \leqslant t,\ S_{n+1} = t + 1 + x\} =$$

$$= \sum_{0\leqslant s\leqslant t} \mathbf{P}\{S_{n+1} = t + 1 + x \mid S_n = s\}\, \mathbf{P}\{S_n = s\} =$$

$$= \sum_{0\leqslant s\leqslant t} \mathbf{P}\{\xi_{n+1} = t + 1 + x - s\}\, \mathbf{P}\{S_n = s\} =$$

$$= \sum_{0\leqslant s\leqslant t} P(t + 1 + x - s)\, \mathbf{P}\{S_n = s\}.$$

Se ve que

$$\mathbf{P}\{\eta(t) = x\} = \sum_{0\leqslant s\leqslant t} P(t - s + 1 + x)\, p_s(0) =$$

$$= \sum_{0\leqslant u\leqslant t} P(u + 1 + x)\, p_{t-u}(0), \tag{5.13}$$

ya que

$$\sum_{n=0}^{\infty} \mathbf{P}\{S_n = s\} = \mathbf{P}\{\xi(s) = 0\} = p_s(0)$$

es la probabilidad de que el momento s sea el momento de restablecimiento. Tomando en consideración que para $t \to \infty$

$$\sum_{u=0}^{t} P(u + 1 + x) \to G(x) \quad \text{y} \quad p_{t-u}(0) \to \frac{1}{\mu}$$

de la igualdad (5.13) obtenemos la fórmula (5.12).

Consideremos ahora el número de restablecimientos $\nu(t+s) - \nu(t)$ durante el intervalo de tiempo $(t, t+s]$. Evidentemente,

$$\mathbf{P}\{\nu(t+s) - \nu(t) = 0\} = \mathbf{P}\{\eta(t) \geqslant s\}, \tag{5.14}$$

y para $n = 1, 2, \ldots$ la probabilidad condicional de que durante el intervalo de tiempo $(t, t+s]$ ocurran n restablecimientos con la condición de que el primer restablecimiento se realiza en el momento $t+u+h$, es

$$\mathbf{P}[\nu(t+s) - \nu(t) = n \mid \eta(t) = u] = \begin{cases} 0, & u \geqslant s \\ \mathbf{P}\{\nu(t+s) - \nu(t+u+1) = n-1\}, & u < s. \end{cases}$$

Utilizando la fórmula de la probabilidad total, obtenemos

$$\mathbf{P}\{\nu(t+s) - \nu(t) = n\} = \sum_{0 \leqslant u < s} \mathbf{P}\{\nu(t+s) - \nu(t+u+1) = n-1\}\, \mathbf{P}\{\eta(t) = u\},$$

que junto con la fórmula (5.14) nos da la relación recurrente para la distribución de probabilidades del número de restablecimientos durante cualquier intervalo $(t, t+s]$. En particular, *para las probabilidades límites*

$$P_n(s) = \lim_{t \to +\infty} \mathbf{P}\{\nu(t+s) - \nu(t) = n\} \tag{5.15}$$

tenemos *las relaciones recurrentes siguientes*:

$$P_0(s) = \frac{1}{\mu} \sum_{u \geqslant s} G(u),$$

$$P_n(s) = \frac{1}{\mu} \sum_{0 \leqslant u < s} P_{n-1}(s-u-1)\, G(u), \quad n = 1, 2, \ldots \tag{5.16}$$

Señalemos que las distribuciones de probabilidades de las magnitudes

$$\xi(t) = t - S_{\nu(t)}, \quad \eta(t) = S_{\nu(t)+1} - t - h, \quad \nu(t+s) - \nu(t) \tag{5.17}$$

coincidirán con las distribuciones estacionarias límites (descritas por las fórmulas (5.7), (5.12), (5.16)), si en el momento inicial $t = 0$, el primer aparato ya ha trabajado un tiempo aleatorio determinado $\xi(0)$, en donde la magnitud aleatoria $\xi(0)$ tiene la distribución estacionaria $\mathbf{P}\{\xi(0) = x\} = G(x)/\mu$, $x = 0, 1, \ldots$

Como conclusión de este punto es necesario decir que los resultados obtenidos anteriormente, también tienen lugar en el caso general[1]). En particular, para los sumandos $\xi_1, \xi_2, \ldots$ distribuidos continuamente con la densidad de probabilidad $p(x)$, las distribuciones de probabilidades de las magnitudes correspondientes

$$\xi(t) = t - S_{\nu(t)}, \quad \eta(t) = S_{\nu(t)+1} - t$$

convergen débilmente hacia la distribución límite con la densidad

$$p^*(x) = \frac{G(x)}{\mu}, \quad 0 \leqslant x < \infty, \tag{5.18}$$

donde

$$G(x) = \int_x^\infty p(y)\,dy, \quad \mu = \int_0^\infty xp(x)\,dx = \int_0^\infty G(x)\,dx. \tag{5.19}$$

2. Sucesión de sumas de las magnitudes independientes. Distribución del máximo. En este punto examinemos algunas propiedades generales de la sucesión de sumas

$$S_n = \sum_{k=1}^{n} \xi_k, \quad n = 1, 2, \ldots, \tag{5.20}$$

de las magnitudes aleatorias independientes ξ_k con igual distribución de probabilidades. Para mayor claridad nos podemos imaginar que una determinada partícula fluctúa aleatoriamente por la recta real, desplazándose a cada paso en una magnitud ξ_n, entonces $S_n = \sum_{k=1}^{n} \xi_k$, $n = 1, 2, \ldots$ ($S_0 = 0$) será la trayectoria de esta fluctuación aleatoria. Supondremos que las magnitudes $\xi_1, \xi_2, \ldots$ tienen una esperanza matemática a distinta de cero, considerando que

$$a = \mathbf{M}\xi_1 < 0.$$

Según la ley reforzada de los grandes números $\frac{S_n}{n} \to a < 0$; por consiguiente, con la probabilidad 1

$$S_n \to -\infty \quad \text{para} \quad n \to \infty,$$

y existe tal ν dependiente del caso, en el cual $S_n < 0$, para todos los valores $n \geqslant \nu$. Esto indica que el máximo

$$\zeta = \text{máx}\,(S_0, S_1, \ldots) \tag{5.21}$$

de la trayectoria de fluctuación aleatoria es una magnitud finita que coincide con el valor máximo de la sucesión $S_0, S_1, \ldots, S_\nu$.

[1]) Referente a esta cuestión (y también en relación con los puntos 2, 3 siguientes) véase por ejemplo, el libro de V. Feller «Introducción a la teoría de probabilidades y sus aplicaciones» («An Introduction to the Probability Theory and its Applications» N. Y., 1950, 1954), tomo II, IL, Moscú, 1967.

Más adelante hallaremos el enlace de la distribución de probabilidades de esta magnitud ζ con un determinado proceso de restablecimiento $S_0^*, S_1^*, \ldots$, construido según la fluctuación aleatoria $S_0, S_1, \ldots$ (véase más adelante (5.31)), lo que permite deducir las fórmulas explícitas (5.37), (5.40) así como también la fórmula (5.41), que se utilizarán al estudiar los procesos del servicio en masa.

Supongamos que

$$\overline{S}_0 = 0, \ \overline{S}_1 = \xi_2, \ \overline{S}_2 = \xi_2 + \xi_3, \ \ldots$$

y

$$\overline{\zeta} = \max (\overline{S}_0, \ \overline{S}_1, \ \ldots).$$

Claro está, que la magnitud aleatoria $\overline{\zeta}$ tiene la misma distribución de probabilidades que la magnitud aleatoria ζ. Evidentemente, para cualquier $z \geqslant 0$, el suceso $\{\overline{\zeta} \leqslant z\}$ se realiza entonces y sólo entonces, cuando $\xi_1 \leqslant z$ y $\overline{\zeta} \leqslant z - \xi_1$. Ya que la magnitud $\overline{\zeta}$ no depende de ξ_1, para cada valor fijado $\xi_1 = x$ $(x \leqslant z)$, tenemos que

$$\mathbf{P}\{\overline{\zeta} \leqslant z - \xi_1 \mid \xi_1\} = F_\zeta (z - \xi_1),$$

donde $F_\zeta (z) = \mathbf{P}\{\zeta \leqslant z\}$ es la función de distribución de la magnitud ζ, y de tal modo

$$\mathbf{P}\{\zeta \leqslant z \mid \xi_1\} = \begin{cases} 0, & \text{si } \xi_1 > z, \\ F_\zeta (z - \xi_1), & \text{si } \xi_1 \leqslant z. \end{cases}$$

Teniendo en cuenta que para el valor ζ no negativo, la función de distribución $F_\zeta (z)$ es igual a cero para $z < 0$, la relación obtenida anteriormente se puede escribir en la forma

$$\mathbf{P}\{\zeta \leqslant z \mid \xi_1\} = F_\zeta (z - \xi_1).$$

Utilizando la fórmula de la probabilidad total (véase (4.20), cap. 1) obtenemos como resultado la siguiente ecuación para la función $F_\zeta (z)$:

$$F_\zeta (z) = \mathbf{M} F_\zeta (z - \xi_1), \quad z \geqslant 0. \tag{5.22}$$

Esta ecuación en el caso de magnitudes discretas $\xi_1, \xi_2, \ldots$ significa que

$$F_\zeta (z) = \sum_{x \leqslant z} F_\zeta (z - x) P_{\xi_1} (x) = \sum_{y \geqslant 0} P_{\xi_1} (z - y) F_\zeta (y), \tag{5.23}$$

y en el caso cuando las magnitudes $\xi_1, \xi_2, \ldots$ tienen la densidad de probabilidad $p_{\xi_1} (x)$,

$$F_\zeta (z) = \int_{-\infty}^{z} F_\zeta (z - x) p_{\xi_1} (x) \, dx = \int_{0}^{\infty} p_{\xi_1} (z - y) F_\zeta (y) \, dy. \tag{5.24}$$

Mostremos que el suceso $\{\zeta = 0\}$ se realiza con la probabilidad positiva $q > 0$:

$$q = F_\zeta(0) > 0 \tag{5.25}$$

(es decir, al salir la partícula desde el punto $x = 0$ con probabilidad $q \neq 0$, nunca cae en el semieje positivo $x > 0$).

En efecto, supongamos lo contrario. Designemos por z_0 la cara inferior de aquellos valores de z, para los cuales $F_\zeta(z) > 0$; de acuerdo con nuestra suposición, será $z_0 > 0$, o bien $z_0 = 0$ y $F_\zeta(0) = 0$. Si $z_0 > 0$, para cualquier z, $0 \leqslant z \leqslant z_0$ de la ecuación (5.22) obtenemos

$$F_\zeta(z) = \mathbf{M} F_\zeta(z - \xi_1) = 0.$$

Evidentemente, en el caso de la magnitud no negativa $F_\zeta(r - \xi_1)$, cuando $F(z - \xi_1) > 0$ para $z - \xi_1 > z_0$, la igualdad indicada puede existir sólo entonces, cuando $z - \xi_1 \leqslant z_0$, $\xi_1 \geqslant z - z_0$ con la probabilidad 1, de donde deducimos, que para $z_0 > 0$ es $\xi_1 \geqslant 0$ con la probabilidad 1. Pero esto contradice nuestra condición inicial $a = \mathbf{M}\xi_1 < 0$, y por consiguiente, deberá ser $z_0 = 0$. Exactamente igual conduce a una contradicción la suposición de que $F_\zeta(0) = 0$ con la condición $\mathbf{M}\xi_1 < 0$, en el caso de $z_0 = 0$.

La magnitud ζ toma rigurosamente un valor positivo con la probabilidad $p = 1 - q$. Esto quiere decir que en el caso de fluctuación aleatoria considerado, la partícula cae tarde o temprano en la parte positiva de la recta real con la probabilidad p; designemos por ξ_1^* su posición en la primera caída en el intervalo $(0, \infty)$. La magnitud ξ_1^* sólo existe con la probabilidad p:

$$\mathbf{P}\{\xi_1^* \in (0, \infty)\} = p \quad (= 1 - q)$$

(consideraremos convencionalmente, que la magnitud ξ_1^* «desaparece» con la probabilidad $q = 1 - p$).

Designemos por τ_1 el momento de caída de la partícula en el intervalo $(0, \infty)$; $\tau_1 = \nu_1^*$, donde ν_1^* es el número de pasos hasta que el desplazamiento de la partícula se haga positiva por primera vez: $S_n - S_0 \leqslant 0$ para $n < \tau_1$, $S_{\tau_1} - S_0 = \xi_1^* > 0$.

Supongamos que $S_1^* = S_{\tau_1}$. El desplazamiento de la partícula se realiza, después del momento τ_1, según la misma ley como después del momento inicial $\tau_0 = 0$. En particular, con la misma probabilidad $p = 1 - q$ que antes, la partícula cae, en un momento determinado τ_2 (después de un número aleatorio de pasos $\nu_2^* = \tau_2 - \tau_1$), en el punto $S_2^* = S_{\tau_2}$, más a la derecha del punto inicial S_1^* (ν_2^* es el número de pasos después del momento τ_1 antes de que el desplazamiento de la partícula se haga de nuevo positivo: $S_n - S_1^* \leqslant 0$ para $n < \tau_2$, $S_{\tau_2} - S_1^* = \xi_2^* < 0$). Empezando desde el momento $\tau_2 = \nu_1^* + \nu_2^*$ se observa el mismo cuadro: con la probabilidad p, en un momento determinado τ_3 (después de un número

aleatorio de pasos $\nu_3^* = \tau_3 - \tau_2$) la partícula cae en el punto $S_3^* = S_{\tau_3}$, más a la derecha del punto inicial S_2^* (ν_2^* es el número de pasos después del momento τ_2 antes de que el desplazamiento de la partícula se haga de nuevo positivo: $S_n - S_2^* \leqslant 0$ para $n \leqslant \tau_3$, $S_{\tau_3} - S_2^* = \xi_3^* > 0$). Continuando este proceso más adelante llegamos a la sucesión de sumas $S_n^* = S_{\tau_n}$,

$$S_n^* = \sum_{k=1}^{n} \xi_k^*, \quad n = 1, 2, \ldots. \tag{5.26}$$

de las magnitudes aleatorias positivas ξ_k^*, $k = 1, 2, \ldots$, donde ξ_{k+1}^* representa en sí el desplazamiento de la partícula (después de su salida desde el punto S_k^*) en el momento τ_{k+1}, cuando este desplazamiento se hace positivo por primera vez: $S_n - S_k^* \leqslant 0$, para $n < \tau_{k+1}$, $S_{\tau_{k+1}} - S_k^* = \xi_{k+1}^* > 0$.

Ya que las magnitudes $\xi_1, \xi_2, \ldots$, son independientes y tienen una distribución igual de probabilidades, el desplazamiento de la partícula después de su salida del punto inicial S_k^* se realiza según la misma ley, que después de la salida del punto $S_0^* = 0$ (independientemente de su comportamiento antes del momento τ_k); por eso *las magnitudes* $\xi_1^*, \xi_2^*, \ldots$ definidas por nosotros, *son independientes y tienen una misma distribución de probabilidades*.

En la terminología del punto 1 anterior, la sucesión S_n^*, $n = 1, 2, \ldots$, *es un proceso de restablecimiento. Con ello, el primer* momento S_1^* de restablecimiento sólo existe con la probabilidad $\mathbf{P}\{S_1^* \in (0, \infty)\} = p$, y hablando en general, al existir los momentos $S_1^*, \ldots, S_{n-1}^*$, el momento S_n^* de restablecimiento siguiente, sólo existe con la misma probabilidad p:

$$\mathbf{P}\{S_{n+1}^* \in (0, \infty) \mid S_n^* \in (0, \infty)\} = p.$$

De aquí obtenemos, para la probabilidad $\mathbf{P}\{S_{n+1}^* \in (0, \infty)$ de existencia del restablecimiento S_{n+1}^* en el momento $n + 1$, que

$$\mathbf{P}\{S_{n+1}^* \in (0, \infty)\} =$$
$$= \mathbf{P}\{S_{n+1}^* \in (0, \infty) \mid S_n^* \in (0, \infty)\}\, \mathbf{P}\{S_n^* \in (0, \infty)\} =$$
$$= p \cdot \mathbf{P}\{S_n^* \in (0, \infty)\}, \quad n = 1, 2, \ldots,$$

y como resultado, tenemos que

$$\mathbf{P}\{S_n^* \in (0, \infty)\} = p^n, \quad n = 1, 2, \ldots. \tag{5.27}$$

Puesto que la serie $\sum_{n=1}^{\infty} \mathbf{P}\{S_n^* \in (0, \infty)\}$ es convergente, según el lema de Borel—Cantelli se realiza con la probabilidad 1, sólo un número finito de sucesos $\{S_n^* \in (0, \infty)\}$, es decir, *el proceso de restablecimiento* S_n^*, $n = 1, 2, \ldots$, *se interrumpe en un determinado paso.* Se ve fácilmente que no habrá ningún restablecimiento con la proba-

bilidad $q = 1 - p$, y sólo habrá un restablecimiento en todo el intervalo infinito $[0, \infty)$ con la probabilidad $p(1-p)$, y en general, el proceso de restablecimiento S_n^*, $n = 1, 2, \ldots$, se interrumpe después de n restablecimientos con la probabilidad $p^n(1-p)$. Suponiendo que $S_0^* = 0$ y considerando también a S_0^*, como momento de restablecimiento obtenemos, para el número total N de restablecimientos (en todo el intervalo infinito $[0, \infty)$) la distribución geométrica de probabilidades:

$$\mathbf{P}\{N \geqslant n\} = \mathbf{P}\{S_{n-1}^* \in (0, \infty)\} = p^{n-1} \qquad (5.28)$$
$$\mathbf{P}\{N = n\} = qp^{n-1}, \quad n = 1, 2, \ldots$$

Designemos por $N(x)$ al número de restablecimientos en el intervalo $[0, x]$, $0 \leqslant x < \infty$. Evidentemente,

$$\mathbf{P}\{N(x) \geqslant n\} = \mathbf{P}\{S_{n-1}^* \in [0, x]\}$$

y el valor medio $M(x) = \mathbf{M}[N(x)]$ es (compárese con (5.5))

$$M(x) = \sum_{n=0}^{\infty} \mathbf{P}\{N(x) > n\} = \sum_{n=0}^{\infty} \mathbf{P}\{S_n^* \in [0, x]\}. \qquad (5.29)$$

Para el número total N de restablecimientos

$$\mathbf{M}[N] = \sum_{n=0}^{\infty} \mathbf{P}\{N > n\} = \sum_{n=0}^{\infty} p^n = \frac{1}{q} \qquad (5.30)$$

Evidentemente, la magnitud $\zeta = \max(0, S_1, S_2, \ldots)$ que nos interesa coincide con el valor máximo de la sucesión monótonamente creciente $0, S_1^*, S_2^*, \ldots$, en otras palabras, con el último momento S_{N-1}^* de restablecimiento. Por eso para cualquier $x > 0$

$$\mathbf{P}\{0 < \zeta \leqslant x\} = \sum_{n>1} \mathbf{P}\{N = n, \quad S_{n-1}^* \in [0, x]\}.$$

Tenemos que

$$\mathbf{P}\{N = n, \ S_{n-1}^* \in [0, x]\} = \mathbf{P}\{S_{n-1}^* \in [0, x] \mid N = n\}\, \mathbf{P}\{N = n\},$$

donde el suceso $\{N = n\}$ significa que $\{S_{n-1}^* \in (0, \infty)\}$, y la magnitud ξ_n^* «desaparece». Pero $S_{n-1}^* = \sum_{k=1}^{n-1} \xi_k^*$ no depende de ξ_n^* ($\xi_1^*, \xi_2^*, \ldots$ es la sucesión de magnitudes i n d e p e n d i e n t e s) y se ve fácilmente que

$$\mathbf{P}\{S_{n-1}^* \in [0, x] \mid N = n\} = \mathbf{P}\{S_{n-1}^* \in [0, x] \mid S_{n-1}^* (0, \infty)\}.$$

De las fórmulas obtenidas anteriormente (5.27), (5.28) deducimos luego que

$$\mathbf{P}\{S_{n-1}^* \in [0, x] \mid S_{n-1}^* \in (0, \infty)\}\, \mathbf{P}\{N = n\} =$$
$$= \frac{\mathbf{P}\{S_{n-1}^* \in [0, x]\}}{\mathbf{P}\{S_{n-1}^* \in (0, \infty)\}}\, \mathbf{P}\{N = n\} = q\mathbf{P}\{S_{n-1}^* \in [0, x]\}, \quad n > 1.$$

Teniendo en cuenta la igualdad (5.29), como resultado obtenemos la siguiente relación:

$$F_\zeta(x) = qM(x) \quad (q = F_\zeta(0)). \tag{5.31}$$

Para la función $M(x)$, o sea, el número medio de restablecimientos en el intervalo $[0, x]$, obtenemos la fórmula (5.29) en la cual

$$\mathbf{P}\{S_0^* \in [0, x]\} = F^{0*}(x) \equiv 1,$$

y para $n > 0$

$$\mathbf{P}\{S_n^* \in [0, x]\} = F^{n*}(x), \quad 0 \leqslant x < \infty,$$

es la función de distribución de la suma de las magnitudes independientes $\xi_1^*, \ldots, \xi_n^*$, de igual distribución. Conociendo la función de distribución $F^*(x) = \mathbf{P}\{\xi_1^* \leqslant x\}$, en principio se puede determinar también que

$$M(x) = \sum_{n=0}^{\infty} F^{n*}(x), \quad 0 \leqslant x < \infty \tag{5.32}$$

($F^{n*}(x) = 0$ cuando $x < 0$ para todos los $n = 0, 1, \ldots$).

Deduzcamos una relación útil que, en el caso de la distribución exponencial de las magnitudes iniciales $\xi_1, \xi_2, \ldots$ permitirá obtener expresiones explícitas para la distribución de probabilidades de las magnitudes ξ_t^* (véase (5.35) y más adelante).

Precisamente, en la fluctuación aleatoria inicial (5.20), la magnitud ξ_t^* significa la posición de la partícula en la primera caída en el intervalo $(0, \infty)$: $\xi_t^* = S_{\tau_1}$, donde τ_1 es el número de pasos hasta la primera caída en el intervalo $(0, \infty)$. Evidentemente, para cualquier $x \geqslant 0$

$$\mathbf{P}\{\tau_1 = n, \ \xi_t^* > x\} = \mathbf{P}\{S_1 \leqslant 0, \ \ldots, \ S_{n-1} \leqslant 0, \ S_n > x\},$$

y para los valores fijados $S_1 \leqslant 0, \ldots, S_{n-1} \leqslant 0$, el suceso $\{\tau_1 = n, \ \xi_t^* < x\}$ significa que $\xi_n = S_n - S_{n-1} > x - S_{n-1}$. Por consiguiente, $\mathbf{P}\{\tau_1 = n, \ \xi_t^* > x \mid S_1, \ \ldots, \ S_{n-1}\} = G_{\xi_1}(x - S_{n-1}) \times \chi_{(S_1 \leqslant 0, \ \ldots, \ S_{n-1} \leqslant 0)}$, donde $G_{\xi_1}(x) = 1 - F_{\xi_1}(x)$, $F_{\xi_1}(x)$ es la función de distribución de las magnitudes $\xi_1, \xi_2, \ldots$, y $\chi_{(S_1 \leqslant 0, \ \ldots, \ S_{n-1} \leqslant 0)}$ es el indicador del suceso[1] $\{S_1 \leqslant 0, \ \ldots, \ S_{n-1} \leqslant 0\}$. Según la fórmula de la probabilidad total

$$\mathbf{P}\{\tau_1 = n, \ \xi_t^* > x\} = \mathbf{M}[G_{\xi_1}(x - S_{n-1}) \cdot \chi_{(S_1 \leqslant 0, \ \ldots, \ S_{n-1} \leqslant 0)}], \tag{5.33}$$
$$n = 1, 2, \ldots; \ x \geqslant 0.$$

[1]) Recordemos que el indicador del suceso A se determina como

$$\chi_A = \begin{cases} 1 \text{ para la aparición de } A, \\ 0 \text{ en caso contrario.} \end{cases}$$

Supongamos que cada una de las magnitudes $\xi_1, \xi_2, \ldots$, está distribuida en el intervalo $(0, \infty)$ según la ley exponencial:

$$G_{\xi_i}(x) = \mathbf{P}\{\xi_i > x\} = Ae^{-\lambda x}, \quad x \geqslant 0, \tag{5.34}$$

donde $A = \mathbf{P}\{\xi_i > 0\}$. En este caso la relación (5.33), en la cual S_{n-1} figura solamente, para la condición $S_{n-1} \leqslant 0$, nos da las siguientes igualdades:

$$\mathbf{P}\{\tau_1 = n,\ \xi_1^* > x\} = \\ = \mathbf{M}[Ae^{\lambda x - \lambda S_{n-1}} \chi_{\{S_1 \leqslant 0, \ldots, S_{n-1} \leqslant 0\}}] = B_n e^{-\lambda x},$$

donde

$$B_n = \mathbf{M}[Ae^{-\lambda S_{n-1}} \chi_{\{S_1 \leqslant 0, \ldots, S_{n-1} \leqslant 0\}}];$$

sumando estas igualdades por todos los valores de $n = 1, 2, \ldots$ obtenemos que

$$\mathbf{P}\{\xi_1^* > x\} = pe^{-\lambda x}, \quad x \geqslant 0, \tag{5.35}$$

donde la constante $p \sum_{n=1}^{\infty} B_n$ es, evidentemente,

$$p = \mathbf{P}\{\xi_1^* > 0\} = 1 - q \quad (q = F_\xi(0)).$$

Examinemos más detalladamente, el caso de las magnitudes distribuidas continuamente $\xi_1, \xi_2, \ldots$ (para las magnitudes discretas los resultados son completamente análogos). De acuerdo con la fórmula (5.35) para la densidad de probabilidad

$$p_{\xi_i}(x) = A\lambda e^{-\lambda x}, \quad x > 0,$$

las magnitudes $\xi_1^*, \xi_2^*, \ldots$ estarán distribuidas según la ley exponencial con la densidad

$$p^*(x) = p\lambda e^{-\lambda x}, \quad x > 0, \tag{5.36}$$

que se diferencia de la densidad exponencial corriente, sólo en el factor p, que indica que la magnitud ξ_1^* «desaparece» con la probabilidad $q = 1 - p$. Como sabemos (véase (2.16), cap. II), la suma de las magnitudes independientes $\xi_1^*, \ldots, \xi_n^*$ con la misma distribución exponencial (5.36) tiene la densidad de probabilidad

$$p^{n*}(x) = p^n \lambda \frac{(\lambda x)^{n-1}}{(n-1)} e^{-\lambda x}, \quad x > 0.$$

La densidad $p^{n*}(x)$ es la derivada de $F^{n*}(x)$ y, como se ve fácilmente,

$$m(x) = \sum_{n=1}^{\infty} p^{n*}(x) = p\lambda e^{-\lambda x} \sum_{n=0}^{\infty} \frac{(p\lambda x)^n}{n!} = p\lambda e^{-\lambda(1-p)x}, \quad x > 0$$

es la derivada de la función $M(x) = \sum_{n=0}^{\infty} F^{n*}(x)$ $(M(0) = 1)$, y, por consiguiente,

$$M(x) = 1 + \int_0^x m(x)\,dx = \frac{1}{q}(1 - pe^{-\lambda q x}), \quad q = 1 - p.$$

Como resultado, de (5.31) obtenemos que *la función de distribución del máximo* $\zeta = \max(0, S_1, S_2, \ldots)$ *tiene la forma*

$$F_\zeta(x) = 1 - (1 - q)\,e^{-\lambda q x}, \quad x \geqslant 0, \tag{5.37}$$

donde la probabilidad $q = F_\zeta(0)$ *puede ser determinada de la ecuación* (5.24):

$$q = \int_{-\infty}^{0} [1 - (1 - q)\,e^{\lambda q x}]\,p_{\xi_1}(x)\,dx. \tag{5.38}$$

En el caso cuando la densidad de probabilidad para $x \leqslant 0$ tiene el mismo tipo exponencial que para $x > 0$, a saber:

$$p_{\xi_1}(x) = B\mu e^{\mu x}, \quad x > 0, \tag{5.39}$$

de la ecuación (5.38) se deduce fácilmente que $q = B\frac{\lambda + \mu}{\lambda} - \frac{\mu}{\lambda}$; teniendo en cuenta la igualdad $B = \mathbf{P}\{\xi_1 \leqslant 0\} = 1 - A$, tenemos

$$q = B - A\frac{\mu}{\lambda}. \tag{5.40}$$

Señalemos que el caso examinado $\mathbf{M}\xi_1 = -\frac{B}{\mu} + \frac{A}{\lambda}$, y se ve directamente, que para $\mathbf{M}\xi_1 < 0$ el valor indicado $q = -\mu\mathbf{M}\xi_1$ es rigurosamente positivo.

Volvamos a examinar el caso general de fluctuación aleatoria (5.20), suponiendo como antes, que $\mathbf{M}\xi_1 < 0$.

Mostremos que *la magnitud* ν_0 *es el número de pasos hasta la primera caída (después de salir de* $S_0 = 0$*) en la zona* $x \leqslant 0$ *y tiene una esperanza matemática finita* $\mathbf{M}\nu_0 < \infty$; *y además,*

$$\mathbf{M}\nu_0 = \frac{1}{q} \quad (q = F_\zeta(0)). \tag{5.41}$$

El hecho de que la magnitud ν_0 es finita con la probabilidad 1, fue señalado desde el mismo comienzo (recuérdese que $S_n \to -\infty$ para $n \to \infty$). Para la deducción de la igualdad (5.41) examinemos los sucesos

$$A_n = \{S_n > 0,\ S_n > S_1,\ \ldots,\ S_n > S_{n-1}\}, \quad n = 1, 2, \ldots$$

Evidentemente cada uno de ellos significa que S_n es el momento de restablecimiento en el proceso de restablecimiento (5.26). Teniendo

en cuenta que cada momento de restablecimiento en (5.26) coincide con un valor determinado S_n, e introduciendo el suceso cierto A_0 que significa que $S_0^* = S_0$ es también un momento de restablecimiento, para el número total N de restablecimientos y el valor medio $\mathbf{M}[N]$ tendremos que $N = \sum_{n=0}^{\infty} \chi_{A_n}$, donde χ_{A_n} es el indicador del suceso A_n y

$$\mathbf{M}[N] = \sum_{n=0}^{\infty} \mathbf{M}\chi_{A_n} = \sum_{n=0}^{\infty} \mathbf{P}(A_n) = \frac{1}{q} \qquad (5.42)$$

(véanse (5.30) y (4.39), cap. I). Refiriéndonos a las magnitudes $S'_k = S_n - S_{n-k}$, $k = 1, \ldots, n$:

$$S'_1 = \xi_1, \; S'_2 = \xi_n + \xi_{n-1}, \; \ldots, \; S'_n = \xi_n + \ldots + \xi_1,$$

que tienen la misma distribución de probabilidades que las magnitudes

$$S_1 = \xi_1, \; S_2 = \xi_1 + \xi_2, \; \ldots, \; S_n = \xi_1 + \ldots + \xi_n,$$

se ve fácilmente que cada uno de los sucesos A_n coincide con el suceso correspondiente $\{S'_1 > 0, \ldots, S'_n > 0\}$, que tiene la misma probabilidad que el suceso

$$B_n = \{S_1 > 0, \ldots, S_n > 0\}, \quad n = 1, 2, \ldots$$

Pero cada uno de los sucesos B_n significa que la partícula no cayó en la zona $x \leqslant 0$ en los primeros n pasos, es decir, $B_n = \{\nu_0 > n\}$. Teniendo en cuenta que $\mathbf{P}(B_n) = \mathbf{P}(A_n)$, $n = 0, 1, \ldots$, de la igualdad (5.42) obtenemos para la esperanza matemática de la magnitud positiva ν_0, que

$$\mathbf{M}\nu_0 = \sum_{n=0}^{\infty} \mathbf{P}\{\nu_0 > n\} = \sum_{n=0}^{\infty} \mathbf{P}(A_n) = \frac{1}{q}.$$

La fórmula (5.41) queda demostrada.

3. Procesos aleatorios en los sistemas con una línea de servicio. Supongamos que en un sistema de servicio determinado se reciben demandas en los momentos aleatorios de tiempo $\tau_1, \tau_2, \ldots$ Supongamos que es imposible la recepción simultánea de distintas demandas y que los intervalos $\xi_1 = \tau_2 - \tau_1$, $\xi_2 = \tau_3 - \tau_2, \ldots$ entre los momentos $\tau_1, \tau_2, \ldots$ son magnitudes aleatorias independientes, que tienen una misma distribución de probabilidades. Supongamos, además de esto, que para el servicio de la demanda n se invierte el tiempo η_n y $\eta_1, \eta_2, \ldots$, son magnitudes aleatorias independientes con una misma distribución de probabilidades, también independien-

tes de los momentos de tiempo $\tau_1, \tau_2, \ldots$ (tenemos en cuenta que η_n es el tiempo «neto» para el servicio de la demanda n, sin contar el tiempo aleatorio de espera $\mathscr{H}_n$ desde el momento de recepción τ_n hasta el comienzo del servicio).

El tiempo total invertido por la demanda n en el sistema de servicio, en nuestras designaciones $\mathscr{H}_n + \eta_b$. Supongamos que si la siguiente demanda $(n + 1)$ se recibe después del tiempo $\xi_n \geqslant \mathscr{H}_n + \eta_n$, entonces ésta encuentra libre al sistema de servicio el cual *inmediatamente empieza a servirla, es decir*, $\mathscr{H}_{n+1} = 0$; y si $\xi_n < \mathscr{H}_n + \eta_n$, entonces en el momento $\tau_{n+1} = \tau_n + \xi_n$ el sistema todavía está ocupado por el servicio de las demandas anteriores, y hasta el comienzo del servicio, la demanda $n + 1$ debe esperar un tiempo $\mathscr{H}_{n+1} = \mathscr{H}_n + \eta_n - \xi_n$.

Ahora nos ocuparemos con la ley del proceso de servicio prolongado y ante todo con la distribución de probabilidades de la magnitud $\mathscr{H}_n$, o sea, con el tiempo de espera del comienzo de servicio para la demanda n (para $n \to \infty$).

Supongamos que

$$\Delta_n = \eta_n - \xi_n; \; S_n = \sum_{k=1}^{n} \Delta_k, \quad n = 1, 2, \ldots \tag{5.43}$$

Como fue indicado, las magnitudes $\mathscr{H}_n$ están ligadas con las magnitudes aleatorias i n d e p e n d i e n t e s Δ_n $(n = 1, 2, \ldots)$ por las siguientes relaciones:

$$\mathscr{H}_1 = 0, \; \mathscr{H}_{n+1} = \begin{cases} 0 & \text{para } \mathscr{H}_n + \Delta_n \leqslant 0, \\ \mathscr{H}_n + \Delta_n & \text{para } \mathscr{H}_n + \Delta_n > 0. \end{cases} \tag{5.44}$$

Comparemos la sucesión $\mathscr{H}_1, \mathscr{H}_2, \ldots$ con la sucesión $S_1, S_2, \ldots$

Designemos por $\nu_0, \nu_1, \ldots$ los valores sucesivos n, para los cuales $\mathscr{H}_{n+1} = 0$, $n \geqslant 1$. El suceso $\nu_0 = 1$ significa $\{S_1 \leqslant 0\}$, y se ve de la relación (5.44) que $\mathscr{H}_{n+1} = S_n$, para $1 \leqslant n < \nu_0 - 1$ y

$$\{\nu_0 = m\} = \{S_1 > 0, \ldots, S_{m-1} > 0, \; S_m \leqslant 0\}, \quad m > 1. \tag{5.45}$$

Por consiguiente, para $1 \leqslant n \leqslant m$, $m = \nu_0$ el enlace entre $\mathscr{H}_{n+1}$ y S_n se puede expresar formalmente por la siguiente relación:

$$\mathscr{H}_{n+1} = S_n - \text{mín}\,(0, S_1, \ldots, S_n). \tag{5.46}$$

Para $m \leqslant n < \nu_1$, $m = \nu_0$ las magnitudes $\mathscr{H}_{n+1}$ recibieron en cada paso n el mismo incremento Δ_n que las magnitudes S_n: $\mathscr{H}_{n+1} - \mathscr{H}_{m+1} = S_n - S_m > 0$, donde

$$\mathscr{H}_{m+1} = 0, \quad S_m = \text{mín}\,(0, S_1, \ldots, S_n).$$

Del mismo modo que en (5.46), $\mathscr{H}_{n+1} = S_n - \text{mín}\,(0, S_1, \ldots, S_n)$. Para $n = \nu_2$ tenemos $\mathscr{H}_n + \Delta_n = (S_{n-1} - S_m) + \Delta_n = S_n - S_m \leqslant 0$ tal que la fórmula (5.46) que nos da $\mathscr{H}_{n+1} = 0$, es

también válida como anteriormente. Ahora ya debe estar claro que esta fórmula tiene lugar para todos los $n = 1, 2, \ldots$

Examinemos las sumas sucesivas $S_1' = \Delta_n$, $S_2' = \Delta_n + \Delta_{n-1}$, $\ldots$ $\ldots$, $S_n' = \Delta_n + \ldots + \Delta_1$ de las mismas magnitudes aleatorias independientes $\Delta_1, \ldots, \Delta_n$ que en (5.43), pero tomadas en orden contrario. Tenemos

$$S_k' = S_n - S_{n-k}, \quad k = 1, \ldots, n, \tag{5.47}$$

y evidentemente,

$$\text{máx } (0, S_1', \ldots, S_n') = S_n - \text{mín } (0, S_1, \ldots, S_n) = \mathscr{H}_n. \tag{5.48}$$

La distribución conjunta $(S_1', \ldots, S_n')$ es igual a la distribución $(S_1, \ldots, S_n)$ y de tal modo, ha sido obtenido el resultado siguiente.

Lema 1. *La distribución de probabilidades de la magnitud $\mathscr{H}_n$ es la misma que la de la magnitud* $\zeta_n = \text{máx } (0, S_1, \ldots, S_n)$.

Sea a el intervalo medio de tiempo entre las demandas que se reciben sucesivamente ($a = \mathbf{M}\xi_1$), y b el tiempo medio de servicio de una demanda aislada ($b = \mathbf{M}\eta_1$).

El valor medio de las magnitudes $\Delta_n = \eta_n - \xi_n$, $n = 1, 2, \ldots$, es igual a $b - a$. Para la condición

$$\mathbf{M}\Delta_1 = b - a < 0,$$

como fue mostrado en el punto 2 anterior, la magnitud $\zeta = \text{máx } (0, S_1, S_2, \ldots)$ es finita con la probabilidad 1.

La sucesión de magnitudes $\zeta_n = \text{máx } (0, S_1, \ldots, S_n)$, $n = 1, 2, \ldots$, crece monótonamente y converge hacia la magnitud $\zeta = \text{máx } (0, S_1, S_2, \ldots)$. Evidentemente, la desigualdad $\zeta \leqslant x$ es equivalente a que $\zeta_n \leqslant x$ para todos los $n = 1, 2, \ldots$ Además de esto, los sucesos $A_n = \{\zeta_n \leqslant x\}$ forman la sucesión monótona: $A_1 \supseteq A_2 \supseteq \ldots$ y debido a la continuidad de la probabilidad para el suceso $A = \{\zeta \leqslant x\}$ que coinciden con la intersección $\bigcap_n A_n$, tenemos $\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n)$.

Esto junto con el lema 1 da el siguiente resultado.

Teorema 1. *Para la condición*

$$\mathbf{M}\xi_1 = a > b = \mathbf{M}\mu_1 \tag{5.49}$$

la distribución de probabilidades de las magnitudes $\mathscr{H}_n$ converge hacia la distribución del máximo $\zeta = (0, S_1, S_2, \ldots)$: para cualquier x

$$\mathbf{P}\{\mathscr{H}_n \leqslant x\} \to F_\zeta(x) \quad \text{para} \quad n \to \infty. \tag{5.50}$$

(Recordemos que la función de distribución $F_\zeta(x)$ del máximo de la sucesión $0, S_1, S_2, \ldots$ de las sumas de las magnitudes aleatorias independientes fue examinada por nosotros en el punto 2 anterior).

Como complemento del teorema 1 señalemos que para la condición

$$\mathbf{M}\Delta_1 = b - a > 0$$

según la ley reforzada de los grandes números $\frac{S_n}{n} \to \mathbf{M}\Delta_1$, $S_n \to \infty$ con la probabilidad 1, y, por consiguiente,

$$\zeta_n = \text{máx}\,(0, S_1, \ldots, S_n) \to \infty. \tag{5.51}$$

Para las magnitudes $\mathscr{H}_n$, que tienen la misma distribución de probabilidades que las magnitudes ζ_n, se deduce de la relación (5.51) que

$$\mathbf{P}\{\mathscr{H}_n \geqslant x\} \to 1 \quad \text{para} \quad n \to \infty \tag{5.52}$$

para cualquier x tan grande como se quiera. Hablando a grueso modo, esto significa que las demandas $\mathscr{H}_n n \to \infty$, recibidas lo suficientemente tarde, con la probabilidad próxima a 1, esperarán su servicio un tiempo infinitamente largo.

Ejemplo. Supongamos que el tiempo de servicio de la demanda de turno tiene una distribución exponencial de probabilidades con la densidad

$$p_{\eta_1}(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

y sea $p_{\xi_1}(y)$ la densidad de probabilidad de la magnitud ξ_1 ($p_{\xi_1}(y) = 0$ para $y < 0$). Entonces la suma $\Delta_1 = \eta_1 - \xi_1$ de las magnitudes independientes η_1 y $-\xi_1$ tiene la densidad

$$p(x) = \int_{-\infty}^{0} p_{\eta_1}(x-y)\, p_{-\xi_1}(y)\, dy = \int_{0}^{\infty} p_{\eta_1}(x+y)\, p_{\xi_1}(y)\, dy,$$

que para $x > 0$ se da por la fórmula

$$p(x) = \int_{0}^{\infty} \lambda e^{-\lambda(x+y)} p_{\xi_1}(y)\, dy = A\lambda e^{-\lambda x}, \quad x > 0,$$

donde

$$A = \int_{0}^{\infty} e^{-\lambda y} p_{\xi_1}(y)\, dy = \mathbf{P}\{\Delta_1 > 0\}.$$

Como establecimos anteriormente, en este caso la función de distribución $F_\zeta(x)$ se describe por la fórmula (5.37), a saber:

$$F_\zeta(x) = 1 - (1-q)\, e^{-\lambda q x}, \quad x \geqslant 0.$$

Si el tiempo de espera de la demanda siguiente también tiene la ley exponencial de distribución

$$p_{\xi_1}(x) = \begin{cases} \mu e^{-\mu x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

entonces la densidad de probabilidad $p(x)$ para $x < 0$ será

$$p(x) = B\mu e^{\mu x},$$

donde $B = \mathbf{P}\{\Delta_1 \leqslant 0\} = 1 - A$, siendo en la densidad indicada $p(x)$

$$A = \frac{\mu}{\lambda+\mu}, \quad B = \frac{\lambda}{\lambda+\mu}.$$

Según la fórmula (5.40), la probabilidad $q = F_\zeta(0)$, para grandes valores de n, es aproximadamente igual a la probabilidad de que la demanda n encuentre libre al sistema, será

$$q = B - A\frac{\mu}{\lambda} = 1 - \frac{\mu}{\lambda}.$$

(Recordemos, que el intervalo medio de tiempo entre los momentos en que se reciben las demandas es $a = \frac{1}{\mu}$, el tiempo medio de servicio de cada demanda aislada es $b = \frac{1}{\lambda}$ y la condición (5.49) en el teorema 1 significa que $\lambda > \mu$.)

Volvamos al esquema general, suponiendo que sólo se satisface la condición (5.49), y examinemos *el período de ocupación* del sistema de servicio, o sea, el intervalo de tiempo desde el momento de recibir la primera demanda hasta el momento, cuando el sistema queda de nuevo libre.

Si ν_0 es el número de todas las demandas, que se sirven en el período de ocupación, entonces la duración de este período será $T = \sum_{k=1}^{\nu_0} \eta_k$. De la relación (5.45) se ve que ν_0 coincide con el número de pasos hasta la primera entrada de la sucesión $S_n = \sum_{k=1}^{n} \Delta_k$, $n = 1, 2, \ldots$ en la zona $x \leqslant 0$. Como fue mostrado en el punto 2 anterior, para la condición (5.49) la magnitud aleatoria ν_0 tiene una esperanza matemática finita $\mathbf{M}\nu_0$ igual a $\frac{1}{q}$ ($q = F_\zeta(0)$, véase la fórmula (5.41)).

Calculemos la esperanza matemática de la magnitud aleatoria

$$T = \sum_{k=1}^{\nu_0} \eta_k.$$

Lema 2. *Tiene lugar la siguiente igualdad:*

$$\mathbf{M}\sum_{k=1}^{\nu_0} \eta_k = \mathbf{M}\eta_1 \mathbf{M}\nu_0 \tag{5.53}$$

se denomina corrientemente *identidad de Wald*).

Demostración. Aquí tratamos sobre la esperanza matemática de la suma de un número aleatorio de magnitudes $\eta_1, \eta_2, \ldots$ igualmente distribuidas. La igualdad (5.53) es evidente para las magnitudes $\eta_1, \eta_2, \ldots$, independientes del número ν_0, ya que para la condición $\nu_0 = n$ $\mathbf{M}\left(\sum_{k=1}^{\nu_0} \eta_k \mid \nu_0 = n\right) = n\mathbf{M}\eta_1$ y según la fórmula de la esperanza matemática total (véase (4.19), cap. I),

$$\mathbf{M}\sum_{k=1}^{\nu_0} \eta_k = \mathbf{M}\eta_1 \sum_{n=1}^{\infty} n\mathbf{P}\{\nu_0 = n\} = \mathbf{M}\eta_1 \cdot \mathbf{M}\nu_0.$$

Utilicemos un método análogo para la deducción de la fórmula (5.53), y en nuestro caso, cuando las magnitudes $\eta_1, \eta_2, \ldots$ y el número ν_0 están ligados con una misma sucesión S_n, $n = 1, 2, \ldots$, y son magnitudes independientes. Tenemos

$$\begin{aligned}
\mathbf{M}\left(\sum_{k=1}^{\nu_0} \eta_k\right) &= \mathbf{M}(\eta_1 \mid \nu_0 \leqslant 1)\mathbf{P}(\nu_0 \leqslant 1) + \\
&+ \mathbf{M}\left(\eta_1 + \sum_{k=2}^{\nu_0} \eta_k \mid \nu_0 > 1\right)\mathbf{P}(\nu_0 > 1) = \\
&= [\mathbf{M}(\eta_1 \mid \nu_0 \leqslant 1)\mathbf{P}(\nu_0 \leqslant 1) + \mathbf{M}(\eta_1 \mid \nu_0 > 1)\mathbf{P}(\nu_0 > 1)] + \\
&+ \mathbf{M}\left(\sum_{k=2}^{\nu_0} \eta_k \mid \nu_0 > 1\right)\mathbf{P}(\nu_0 > 1) = \\
&= \mathbf{M}\eta_1 + \mathbf{M}\left(\sum_{k=2}^{\nu_0} \eta_k \mid \nu_0 > 1\right)\mathbf{P}(\nu_0 > 1).
\end{aligned}$$

El suceso $\{\nu_0 > 1\}$ se determina completamente por las magnitudes η_1, ξ_1, y, por consiguiente, la magnitud η_2 no depende de este suceso (recordemos que las magnitudes aleatorias $\eta_1, \eta_2, \ldots$ son independientes entre sí y de las magnitudes $\xi_1, \xi_2, \ldots$), en particular, $\mathbf{M}(\eta_2 \mid \nu_0 > 1) = \mathbf{M}\eta_2$. Por eso, teniendo en cuenta las distribuciones de probabilidades para la condición $\nu_0 > 1$, así como anteriormente, obtenemos

$$\begin{aligned}
\mathbf{M}\left(\sum_{k=2}^{\nu_0} \eta_k \mid \nu_0 > 1\right) &= \mathbf{M}(\eta_2 \mid \nu_0 > 1) + \\
&+ \mathbf{M}\left(\sum_{k=3}^{\nu_0} \eta_k \mid \nu_0 > 2\right)\mathbf{P}\{\nu_0 > 2 \mid \nu_0 > 1\} = \\
&= \mathbf{M}\eta_2 + \mathbf{M}\left(\sum_{k=3}^{\nu_0} \eta_k \mid \nu_0 > 2\right)\frac{\mathbf{P}(\nu_0 > 2)}{\mathbf{P}(\nu_0 > 1)}
\end{aligned}$$

y, por consiguiente,

$$\mathbf{M}\Big(\sum_{k=2}^{\nu_0}\eta_k \mid \nu_0 > 1\Big)\,\mathbf{P}(\nu_0 > 1) =$$

$$= \mathbf{M}\eta_2 \mathbf{P}(\nu_0 > 1) + \mathbf{M}\Big(\sum_{k=3}^{\nu_0}\eta_k \mid \nu_0 > 2\Big)\,\mathbf{P}(\nu_0 > 2).$$

Utilizando sucesivamente este método para los restantes valores de $n = 3, 4, \ldots$ obtenemos que

$$\mathbf{M}\Big(\sum_{k=n}^{\nu_0}\eta_k \mid \nu_0 > n-1\Big) =$$

$$= \mathbf{M}\eta_n + \mathbf{M}\Big(\sum_{k=n+1}^{\nu_0}\eta_k \mid \nu_0 > n\Big)\frac{\mathbf{P}(\nu_0 > n)}{\mathbf{P}(\nu_0 > n-1)}, \quad n = 3, 4, \ldots,$$

y en resumen,

$$\mathbf{M}\Big(\sum_{k=1}^{\nu_0}\eta_k\Big) = \mathbf{M}\eta_1 + \mathbf{M}\eta_2 \mathbf{P}\{\nu_0 > 1\} + \ldots +$$

$$+ \mathbf{M}\eta_n \mathbf{P}\{\nu_0 > n-1\} + \ldots = \mathbf{M}\eta_1 \sum_{n=0}^{\infty}\mathbf{P}\{\nu_0 > n\} = \mathbf{M}\eta_1 \mathbf{M}\nu_0,$$

que es lo que se exigía demostrar.

La identidad de Wald de la siguiente expresión para el valor medio del período T de ocupación:

$$MT = \frac{b}{q}, \tag{5.54}$$

donde $q = F_\xi(0)$ para grandes valores de n es, aproximadamente, la probabilidad de que la demanda n encuentre libre al sistema $\left(\frac{1}{q} = \mathbf{M}\nu_0\right)$ y $b = \mathbf{M}\eta_1$ es el tiempo medio de servicio de una sola demanda.

Señalemos que la fórmula (5.54) no sólo da el valor medio del primer período de ocupación (desde el momento de recibir la primera demanda), sino también, en general, de cualquier período de ocupación

$$T_1 = \sum_{k=\nu_0+1}^{\nu_1}\eta_k, \qquad T_2 = \sum_{k=\nu_1+1}^{\nu_2}\eta_k, \ \ldots,$$

donde $\nu_0, \nu_1, \ldots$ son los valores sucesivos de n cuando la demanda $n+1$ encuentre libre al sistema.

Examinemos «la duración de turno» $x(t)$ después del tiempo t desde el comienzo del servicio de la primera demanda, que se recibe

en un momento τ_1 (más exacto, $x(t)$ es el número de demandas que se recibieron en el sistema y todavía no han sido servidas). Pongamos

$$\mathbf{P}_k(t) = \mathbf{P}\{x(t) = k\}, \quad k = 0, 1, \ldots;$$

$P_0(t)$ es la probabilidad de que el sistema esté libre en el momento de tiempo $\tau_1 + t$, $P_1(t)$ es la probabilidad de que el sistema tiene una sola demanda que se sirve directamente y así sucesivamente.

Antes observamos que las magnitudes $\mathscr{H}_n$, o sea, el tiempo de espera hasta el comienzo del servicio de la demanda n, tienen una distribución límite de probabilidades cuando $n \to \infty$. Mostremos que la distribución límite para $t \to \infty$ tiene «la duración de turno» $x(t)$.

Designemos por

$$\tau_0^* = \tau_1, \quad \tau_1^* = \tau_1 + \sum_{k=1}^{\nu_0} \xi_k, \quad \tau_2^* = \tau_1^* + \sum_{k=\nu_0+1}^{\nu_1} \xi_k, \ldots$$

a los momentos sucesivos de tiempo, en los cuales las demandas recibidas encuentran libre al sistema (aquí $\nu_0, \nu_1, \ldots$ son los mismos valores que antes, véase la pág. 192). Introduzcamos el siguiente proceso de restablecimiento S_n^*, $n = 1, 2, \ldots$:

$$S_0^* = 0, \quad S_1^* = \tau_1^* - \tau_0^*, \quad S_2^* = \tau_2^* - \tau_1^*, \ldots$$

donde $S_n^* = \sum_{k=1}^{n} \xi_k^*$ *son las sumas de las magnitudes positivas independientes igualmente distribuidas* $\xi_k^* = \tau_k^* - \tau_{k-1}^*$, $k = 1, 2, \ldots$ De acuerdo con la identidad de Wald existe la esperanza matemática $\mathbf{M}\xi_1^*$:

$$\mathbf{M}\xi_1^* = \mathbf{M} \sum_{k=1}^{\nu_0} \xi_k = \mathbf{M}\xi_1 \cdot \mathbf{M}\nu_0 = \frac{a}{q} \tag{5.55}$$

(compárese (5.54)).

Como sabemos, para las magnitudes $\xi^*(t) = t - S^*_{\nu^*(t)}$, donde $\nu^*(t)$ es el número de restablecimientos en el tiempo t, tiene lugar el teorema del restablecimiento: *la distribución de probabilidades de la magnitud* $\xi^*(t)$ *para* $t \to \infty$ *converge débilmente hacia la distribución de cierta magnitud* ξ^*.

Teorema 2. *Supongamos que las probabilidades* $P_k(t)$, $k = 0, 1, \ldots$, *son funciones continuas de* t[1]). *Entonces existen los valores límites*

$$P_k = \lim_{t \to \infty} P_k(t), \quad k = 0, 1, \ldots \tag{5.56}$$

Demostración. Evidentemente, después de cada momento τ_n^*, cuando la demanda que se recibe encuentra libre al sistema, el proceso de formación del turno empieza de nuevo, y por eso, la pro-

[1]) Señalemos que la función de los valores discretos $t = 0, 1, 2, \ldots$ siempre es formalmente continua.

babilidad de que el turno sea igual a k durante el tiempo u, después del momento $\tau_0^* + S_n^*$, es la misma que para $n = 0$, es decir, es igual a $P_k(u)$. Por consiguiente, para el valor fijado $S^*_{\nu^*(t)} = s$, la probabilidad de que $x(t) = k$, será

$$\mathbf{P}\{x(t) = k \mid S^*_{\nu^*(t)}\} = P_k(t - S^*_{\nu^*(t)}) = P_k(\xi^*(t))$$

y

$$P_k(t) = \mathbf{M}P_k(\xi^*(t)).$$

Ahora bien, según la suposición, las distribuciones de las magnitudes aleatorias $\xi^*(t)$ para $t \to \infty$ convergen débilmente hacia la distribución de probabilidades de una magnitud aleatoria ξ^* y de acuerdo al teorema 1 del § 5, cap. II,

$$\mathbf{M}u(\xi^*(t)) \to \mathbf{M}u(\xi^*)$$

para cualquier función continua limitada $u(x)$; para $u(x) = P_k(x)$ obtenemos la relación límite (5.55) en la cual $P_k = \mathbf{M}P_k(\xi^*)$.

§ 6. PROCESOS ALEATORIOS EN LOS SISTEMAS LINEALES

1. Algunas observaciones auxiliares. A un determinado sistema físico lo llamemos lineal, si bajo la acción de las perturbaciones externas $y(s)$, $0 \leqslant s \leqslant t$, su estado fásico en el momento de tiempo t (para las condiciones iniciales nulas) es

$$x(t) = \int_0^t w(t, s)\, y(s)\, ds, \tag{6.1}$$

donde $w(t, s)$, $0 \leqslant s \leqslant t$, es una función continua a trozos: $w(t, s)$, $t \geqslant s$, que corrientemente se llama *función de peso*. Esta describe en el momento de tiempo s, el comportamiento del sistema sacado del estado de reposo por «el impulso unitario» $\delta(t - s)$. (Aquí y en adelante $\delta(t)$ significa «la función delta» por todos conocida).

Por ejemplo, un sistema es lineal, si se describe por la diferencial lineal

$$x^{(n)}(t) + a_1 x^{(n-1)}(t) + \ldots + a_n x(t) = y(t), \tag{6.2}$$

donde los coeficientes $a_k = a_k(t)$, $k = 1, \ldots, n$, pueden cambiar en el transcurso del tiempo t. En este caso la función de peso $w(t, s)$, $t \geqslant s$, para cada valor fijado de s es una solución de la ecuación homogénea

$$\frac{d^n}{dt^n} w(t, s) + a_1 \frac{d^{n-1}}{dt^{n-1}} w(t, s) + \ldots + a_n w(t, s) = 0$$

con las condiciones iniciales

$$\frac{d^k}{dt^k} w(t, s)\Big|_{t=s} = 0, \quad k = 0, 1, \ldots, n-2; \quad \frac{d^{n-1}}{dt^{n-1}} w(t, s)\Big|_{t=s} = 1.$$

(Se puede comprobar directamente, que $x(t) = \int_0^t w(t, s)\, y(s)\, ds$ es una solución de la ecuación diferencial (6.2) que satisface a las condiciones iniciales nulas $x^{(k)}(0) = 0$ para $k = 0, 1, \ldots, n-2$ y $x^{(n-1)}(0) = 1$.)

Examinemos el comportamiento del sistema lineal con la función de peso $w(t, s)$ bajo la acción de las alteraciones caóticas que cambian rápidamente.

Consideremos que sobre el sistema actúan los impulsos aleatorios independientes de la forma $\Delta\eta(t)\,\delta(t - t_k)$, que se originan en los momentos de tiempo indicados t_k, donde $\Delta\eta(t_k)$, $k = 1, 2, \ldots$, son magnitudes aleatorias independientes, que tienen esperanzas matemáticas y dispersiones

$$\mathbf{M}\Delta\eta(t_k) = a_k \Delta t_k, \quad \mathbf{D}\Delta\eta(t_k) = b_k \Delta t_k \tag{6.3}$$

($\Delta t_k = t_k - t_{k-1}$ significa el intervalo de tiempo hasta la aparición del impulso de turno). Como resultado de tal acción «a la salida» del sistema se tendrá el proceso aleatorio $\xi = \xi(t)$, $t \geqslant t_0$ ($t_0 = 0$) de la forma

$$\xi(t) = \sum_{0 \leqslant t_k \leqslant t} w(t, t_k)\, \Delta\eta(t_k), \tag{6.4}$$

para el cual

$$\mathbf{M}\xi(t) = \sum_{0 \leqslant t_k \leqslant t} w(t, t_k)\, \mathbf{M}\Delta\eta(t_k) = \sum_{0 \leqslant t_k \leqslant t} w(t, t_k)\, a_k \Delta t_k,$$

$$\mathbf{D}\xi(t) = \sum_{0 \leqslant t_k \leqslant t} w^2(t, t_k)\, \mathbf{D}\Delta\eta(t_k) = \sum_{0 \leqslant t_k \leqslant t} w^2(t, t_k)\, b_k \Delta t_k.$$

Imaginémonos ¿qué ocurrirá si los intervalos entre los distintos impulsos son muy pequeños? o más exactamente ¿qué ocurrirá cuando máx $\Delta t_k \to 0$?

Para subrayar el enlace de los coeficientes a_k y b_k en la fórmula (6.3) con el intervalo de tiempo correspondiente (t_{k-1}, t_k), introduzcamos las funciones $a(t)$ y $b(t)$ de la forma $a(t) = a_k$, $b(t) = b_k$ para $t_{k-1} < t \leqslant t_k$, $k = 1, 2, \ldots$. Supongamos que estas funciones tienen límite cuando máx $\Delta t_k \to 0$; para sencillez podemos considerar desde el mismo comienzo que

$$a_k = a(t_k), \quad b_k = b(t_k),$$

donde $a(t)$ y $b(t)$ son unas funciones continuas a trozos. Entonces el estado de nuestro sistema en el límite para máx $\Delta t_k \to 0$, en el

momento de tiempo t, será una magnitud aleatoria determinada $\xi(t)$ con la esperanza matemática

$$\mathbf{M}\xi(t) = \lim \sum w(t, t_k)\, a(t_k)\, \Delta t_k = \int_0^t w(t, s)\, a(s)\, ds$$

y con la dispersión

$$\mathbf{D}\xi(t) = \lim \sum w^2(t, t_k)\, b(t_k)\, \Delta t_k = \int_0^t w^2(t, s)\, b(s)\, ds.$$

De las suposiciones que hicimos, se deduce, que las magnitudes $\Delta\eta(t_k)$ son infinitamente pequeñas para máx $\Delta t_k \to 0$, o más exactamente,

$$\mathbf{M}\,[\Delta\eta(t_k)]^2 = b(t_k)\,\Delta t_k + [a(t_k)\,\Delta t_k]^2 \to 0$$

y con ello

$$\sum_{0 \leqslant t_k \leqslant t} \mathbf{D}\Delta\eta(t_k) = \sum b(t_k)\,\Delta t_k \to \int_0^t b(s)\, ds.$$

Si suponemos complementariamente que $[\Delta\eta(t_k)]^3$ tienen en determinado sentido un orden infinitesimal más elevado, más exacto, suponemos el cumplimiento de la condición

$$\sum_{0 \leqslant t_k \leqslant t} \mathbf{M}\,|\Delta\eta(t_k)|^3 \to 0, \tag{6.5}$$

entonces, de acuerdo al teorema límite central, se puede afirmar que en el límite (para máx $\Delta t_k \to 0$) la magnitud $\xi(t)$ tendrá la distribución de probabilidades de Gauss:

$$\mathbf{P}\{x' \leqslant \xi(t) \leqslant x''\} = \frac{1}{\sqrt{2\pi B}} \int_{x'}^{x''} e^{-\frac{(x-A)^2}{2B}}\, dx, \tag{6.6}$$

donde

$$A = \int_0^t w(t, s)\, a(s)\, ds, \quad B = \int_0^t w^2(t, s)\, b(s)\, ds.$$

En efecto, como se ve fácilmente, para la suma de los sumandos independientes $\xi(t) = \sum_{0 \leqslant t_k \leqslant t} w(t, t_k)\,\Delta\eta(t_k)$ se cumplirá la condición conocida de Liapunov:

$$\sum_{0 \leqslant t_k \leqslant t} \mathbf{M}\,|w(t, t_k)\,\Delta\eta(t_k) - \mathbf{M}\,w(t, t_k)\,\Delta\eta(t_k)|^3 \leqslant$$

$$\leqslant \max_{0 \leqslant s \leqslant t} |w(t, s)|^3 \sum \mathbf{M}\,|\Delta\eta(t_k) - a(t_k)\,\Delta t_k|^3 \to 0$$

(recordemos aquí que $\mathbf{M}\xi(t) \to A$, $\mathbf{D}\xi(t) \to B$).

La acción exterior

$$\dot{\eta}(t) = \sum_{0 \leqslant t_k \leqslant t} \Delta\eta(t_k)\,\delta(t - t_k) \tag{6.7}$$

se describe cómodamente, recurriendo al proceso casual

$$\eta(t) = \int_0^t \dot{\eta}(s)\,ds = \sum_{0 \leqslant t_k \leqslant t} \Delta\eta(t_k), \tag{6.8}$$

que hablando a grueso modo, caracteriza la acción total sobre el sistema durante el tiempo t. Partiendo del proceso aleatorio $\eta(t)$, $t \geqslant 0$, se podría representar la magnitud $\Delta\eta(t_k)$ como

$$\Delta\eta(t_k) = \eta(t_k) - \eta(t_{k-1}), \quad k = 1, 2, \ldots,$$

tomando en calidad de $\eta(t)$, $t \geqslant 0$ el proceso correspondiente con incrementos independientes, es decir, tal proceso cuyos incrementos $\Delta\eta(t_k) = \eta(t_k) - \eta(t_{k-1})$, $k = 1, 2, \ldots$, son magnitudes aleatorias independientes (cualesquiera que fuesen $t_0 \leqslant t_1 \leqslant t_2 \leqslant \ldots$). Esta propiedad la poseen, por ejemplo, el proceso del movimiento browniano y el proceso de Poisson ya conocidos.

Para máx $\Delta t_k \to 0$ tratamos de hecho hablando a grueso modo, de la acción continua sobre el sistema con pequeñas perturbaciones independientes, y en esta situación es natural pasar del modelo «discreto» (6.4) al modelo «continuo» correspondiente para el cual puede servir la integral estocástica

$$\xi(t) = \int_0^t w(t, s)\,d\eta(s) \tag{6.9}$$

(más adelante serán dadas su definición y propiedades).

2. Integral estocástica. El proceso aleatorio $\eta = \eta(t)$ se denomina *proceso con incrementos no correlacionados*, si las magnitudes $\eta_1 = \eta(t_1) - \eta(s_1)$ y $\eta_2 = \eta(t_2) - \eta(s_2)$ son no correlacionadas, para cualesquiera $s_1 \leqslant t_1 \leqslant s_2 \leqslant t_2$:

$$\mathbf{M}(\eta_1 - \mathbf{M}\eta_1)(\eta_2 - \mathbf{M}\eta_2) = 0.$$

A este tipo se refiere, naturalmente, cualquier proceso con incrementos independientes (que tienen esperanzas matemáticas y dispersiones finitas) y en particular, el proceso del movimiento browniano y el proceso de Poisson.

En lo sucesivo (véase el § 7) tendremos que examinar los procesos aleatorios c o m p l e j o s.

Señalemos, que para la magnitud compleja η, la *dispersión* se determina como

$$\mathbf{D}\eta = \mathbf{M}\,|\,\eta - \mathbf{M}\eta\,|^2,$$

y las magnitudes aleatorias η_1 y η_2 que toman valores complejos se denominan *no correlacionadas*, si

$$\mathbf{M}[\eta_1 - \mathbf{M}\eta_1]\,\overline{[\eta_2 - \mathbf{M}\eta_2]} = 0. \tag{6.10}$$

Correspondientemente a esto, el proceso aleatorio $\eta(t)$ se llama *proceso con incrementos no correlacionados*, si para cualesquiera incrementos $\eta_1 = \eta(t_1) - \eta(s_1)$, $\eta_2 = \eta(t_2) - \eta(s_2)$ donde $s_1 \leqslant t_1 \leqslant s_2 \leqslant t_2$, se cumple la condición (6.10).

Examinemos tal proceso aleatorio $\eta(t)$, $c_1 \leqslant t \leqslant c_2$, con incrementos no correlacionados, que

$$\mathbf{M}[\eta(t) - \eta(s)] = \int_s^t a(u)\,du, \quad \mathbf{D}[\eta(t) - \eta(s)] = \int_s^t b(u)\,du \tag{6.11}$$

(la función no negativa $b(t)$, $c_1 \leqslant t \leqslant c_2$, en la relación (6.11) la llamamos *densidad de estructura* del proceso aleatorio $\eta(t)$).

Examinemos al principio el caso, cuando $a(t) \equiv 0$.

Para cualquier función constante a trozos $\varphi(t)$, $c_1 \leqslant t \leqslant c_2$, que conserva los valores constantes $y_k = \varphi(t)$, para $t_{k-1} < t \leqslant t_k$, (donde los puntos $c_1 = t_0, t_1, \ldots, t_n = c_2$ dividen el segmento $[c_1, c_2]$ en un número finito de intervalos), la *integral estocástica* se determina por la fórmula

$$\int_{c_1}^{c_2} \varphi(t)\,d\eta(t) = \sum_{k=1}^{n} y_k \Delta\eta(t_k), \tag{6.12}$$

poniendo $\Delta\eta(t_k) = \eta(t_k) - \eta(t_{k-1})$, $k = 1, \ldots, n$. Está claro, que la magnitud $\int_{c_1}^{c_2} \varphi(t)\,d\eta(t)$ no depende del método elegido de división en "intervalos de constancia" $(t_{k-1}t_k)$, $k = 1, \ldots, n$.

Evidentemente, para cualesquiera funciones constante a trozos $\varphi(t)$, $\psi(t)$ y de las constantes λ_1, λ_2,

$$\int_{c_1}^{c_2} [\lambda_1\varphi(t) + \lambda_2\psi(t)]\,d\eta(t) = \lambda_1 \int_{c_1}^{c_2} \varphi(t)\,d\eta(t) + \lambda_2 \int_{c_1}^{c_2} \psi(t)\,d\eta(t) \tag{6.13}$$

Se comprueba fácilmente que

$$\mathbf{M}\int_{c_1}^{c_2} \varphi(t)\,d\eta(t) = 0$$

y como se deduce de la condición (6.10)

$$\mathbf{M}\Big[\int_{c_1}^{c_2} \varphi(t)\,d\eta(t)\,\overline{\int_{c_1}^{c_2} \psi(t)\,d\eta(t)}\Big] = \int_{c_1}^{c_2} \varphi(t)\,\overline{\psi(t)}\,b(t)\,dt \tag{6.14}$$

(se comprueban lo más sencillamente (6.13) y (6.14), eligiendo para φ y ψ la división común $c_1 = t_0, t_1, \ldots, t_n = c_2$).

En particular,

$$\mathbf{M}\left|\int_{c_1}^{c_2}\varphi(t)\,d\eta(t)\right|^2 = \left\|\int_{c_1}^{c_2}\varphi(t)\,d\eta(t)\right\|^2 = \int_{c_1}^{c_2}|\varphi(t)|^2\,b(t)\,dt, \quad (6.15)$$

y para la distancia media cuadrática entre las magnitudes

$$\xi = \int_{c_1}^{c_2}\varphi(t)\,d\eta(t) \quad \text{y} \quad \eta = \int_{c_1}^{c_2}\psi(t)\,d\eta(t)$$

obtenemos la siguiente expresión:

$$\|\xi - \eta\|^2 = \int_{c_1}^{c_2}|\varphi(t) - \psi(t)|^2\,b(t)\,dt. \quad (6.16)$$

Sea $\varphi(t)$, $c_1 \leqslant t \leqslant c_2$, una función que satisface la condición

$$\int_{c_1}^{c_2}|\varphi(t)|^2\,b(t)\,dt < \infty, \quad (6.17)$$

tal que se puede hallar la sucesión de funciones constantes a trozos $\varphi_n(t)$, $n = 1, 2, \ldots$, que convergen hacia $\varphi(t)$ en su media cuadrática:

$$\int_{c_1}^{c_2}|\varphi_n(t) - \varphi(t)|^2\,b(t)\,dt \to 0 \text{ para } n \to \infty \quad (6.18)$$

(siendo también $\int_{c_1}^{c_2}|\varphi_n(t) - \varphi_m(t)|^2\,b(t)\,dt \to 0$, cuando $n,\ m \to \infty$).

La sucesión correspondiente de integrales estocásticas

$$\xi_n = \int_{c_1}^{c_2}\varphi_n(t)\,d\eta(t), \quad n = 1, 2, \ldots,$$

satisfacerá la condición

$$\|\xi_n - \xi_m\|^2 = \int_{c_1}^{c_2}|\varphi_n(t) - \varphi_m(t)|^2\,b(t)\,dt \to 0 \text{ para } n, m \to \infty,$$

y, por consiguiente, existe la magnitud límite ξ de esta sucesión (véase el teorema 2 del § 4, cap. I):

$$\|\xi_n - \xi\| \to 0 \text{ para } n \to \infty.$$

A este límite lo designamos por $\xi = \int_{c_1}^{c_2} \varphi(t)\, d\eta(t)$ y lo llamamos *integral estocástica*[1]) (de la función $\varphi(t)$). Está claro, que la magnitud ξ está determinada unívocamente con la probabilidad 1 y no depende de la elección de la sucesión $\varphi_n(t)$, $n = 1, 2, \ldots$ que converge hacia $\varphi(t)$.

Se ve fácilmete, que para la integral estocástica definida por nosotros se cumplen las relaciones (6.13)—(6.16) indicadas anteriormente, en el caso de las funciones constantes, a trozos $\varphi(t)$ y $\psi(t)$, ya que estas relaciones se conservan en el paso límite utilizado antes, desde las funciones constantes a trozos a las funciones de tipo general que satisfacen la condición (6.17).

Es útil señalar que para las funciones c o n t i n u a s a t r o z o s

$$\int_{c_1}^{c_2} \varphi(t)\, d\eta(t) = \operatorname*{l.i.m}_{n\to\infty} \sum_{k=1}^{n} \varphi(t_{kn})\, \Delta\eta(t_{kn}), \tag{6.19}$$

donde en la parte derecha se toman «las sumas integrales» para una división cada vez más pequeña del intervalo $[c_1, c_2]$ con los puntos t_{kn}, $k = 1, \ldots, n$, lo que corresponde a la definición general de la integral estocástica dada anteriormente, al elegir la sucesión de funciones constantes a trozos $\varphi_n(t)$, $n = 1, 2, \ldots$, del tipo

$$\varphi_n(t) = \varphi(t_{kn}), \quad t_{k-1,n} < t < t_{k,n}; \quad k = 1, \ldots, n.$$

Extendamos la definición de integral estocástica al intervalo infinito $[c_1, c_2]$.

Para cualquier función $\varphi(t)$, $c_1 \leqslant t \leqslant c_2$, que satisface la condición (6.17), eligiendo la sucesión monótona decreciente $c_{1n} \to c_1$ y la sucesión monótona creciente $c_{2n} \to c_2$, $n = 1, 2, \ldots$, para $n, m \to \infty$, tenemos

$$\left\| \int_{c_{1n}}^{c_{2n}} \varphi(t)\, d\eta(t) - \int_{c_{1m}}^{c_{2m}} \varphi(t)\, d\eta(t) \right\|^2 =$$

$$= \int_{c_{1n}}^{c_{1m}} |\varphi(n)|^2 b(t)\, dt + \int_{c_{2m}}^{c_{2n}} |\varphi(t)|^2 b(t)\, dt \to 0$$

[1]) Llamemos la atención sobre la circunstancia de que la «diferencial» $d\eta(t)$ no se puede interpretar en el mismo sentido que para las funciones diferenciables corrientes (para las cuales $d\eta(t) = \eta'(t)\, dt$), ya que el incremento $\Delta\eta(t) = \eta(t + \Delta t) - \eta(t)$ tiene como media el orden $\sqrt{\Delta t}$ (y no Δt como debía ser para la función diferenciable $\eta(t)$; a saber:

$$\mathbf{M}\, |\Delta\eta(t)|^2 \sim b(t)\, \Delta t \quad \text{y} \quad \frac{|\Delta\eta(t)|}{\Delta t} \to \infty \quad \text{para} \quad \Delta t \to 0.$$

(donde para concretizar consideremos $n \geqslant m$), de aquí se deduce que existe el límite

$$\text{l.i.m}_{c_{1n} \to c_1,\ c_{2n} \to c_2} \int_{c_{1n}}^{c_{2n}} \varphi(t)\, d\eta(t) = \int_{c_1}^{c_2} \varphi(t)\, d\eta(t). \tag{6.20}$$

Extendamos ahora el concepto de integral estocástica al caso en que el proceso con incrementos no correlacionados $\eta(t)$, $c_1 \leqslant t \leqslant c_2$, tiene una esperanza matemática distinta de cero

$$\mathbf{M}\eta(t) = \int_{c_1}^{t} a(u)\, du.$$

Para cualquier función $\varphi(t)$, $c_1 \leqslant t \leqslant c_2$ que satisface a la condición

$$\int_{c_1}^{c_2} |\varphi(t)\, a(t)|\, dt < \infty, \qquad \int_{c_1}^{c_2} |\varphi(t)|^2 b(t)\, dt < \infty \tag{6.21}$$

pongamos que

$$\int_{c_1}^{c_2} \varphi(t)\, d\eta(t) = \int_{c_1}^{c_2} \varphi(t)\, d\eta^0(t) + \int_{c_1}^{c_2} \varphi(t)\, a(t)\, dt, \tag{6.22}$$

donde

$$\eta^0(t) = \eta(t) - \mathbf{M}\eta(t), \quad c_1 \leqslant t \leqslant c_2,$$

es un proceso aleatorio con incrementos no correlacionados, del tipo examinado anteriormente (que tiene una esperanza matemática nula) con la densidad de estructura $b(t)$, $c_1 \leqslant t \leqslant c_2$.

Las fórmulas fundamentales (véase (6.14)) para las integrales definidas anteriormente variarán evidentemente de la siguiente forma:

$$\left.\begin{aligned} &\mathbf{M}\Big[\int_{c_1}^{c_2} \varphi(t)\, d\eta(t)\Big] = \int_{c_1}^{c_2} \varphi(t)\, a(t)\, dt, \\ &\mathbf{M}\Big[\int_{c_1}^{c_2} \varphi(t)\, d\eta(t) \cdot \overline{\int_{c_1}^{c_2} \psi(t)\, d\eta(t)}\Big] = \\ &= \int_{c_1}^{c_2} \varphi(t)\, \overline{\psi(t)}\, b(t)\, dt + \int_{c_1}^{c_2} \varphi(t)\, a(t)\, dt \cdot \overline{\int_{c_1}^{c_2} \psi(t)\, a(t)\, dt}; \end{aligned}\right\} \tag{6.23}$$

en lugar de (6.15) tenemos

$$\mathbf{M}\Big|\int_{c_1}^{c_2} \varphi(t)\, d\eta(t)\Big|^2 = \int_{c_1}^{c_2} |\varphi(t)|^2 b(t)\, dt + \Big|\int_{c_1}^{c_2} \varphi(t)\, a(t)\, dt\Big|^2. \tag{6.24}$$

Lo mismo que antes, si la sucesión de funciones $\varphi_n(t)$, $n = 1, 2, \ldots$, es tal que $\int_{c_1}^{c_2} |\varphi_n(t) - \varphi(t)|^2 b(t) \to 0$, y, además de esto

$$\int_{c_1}^{c_2} \varphi_n(t)\, a(t)\, dt \to \int_{c_1}^{c_2} \varphi(t)\, a(t)\, dt,$$

entonces tendremos

$$\int_{c_1}^{c_2} \varphi(t)\, d\eta(t) = \operatorname*{l.i.m}_{n\to\infty} \int_{c_1}^{c_2} \varphi_n(t)\, d\eta(t) \tag{6.25}$$

en el sentido de convergencia de la media cuadrática; en particular, para la función continua a trozos $\varphi(t)$, la integral estocástica $\int_{c_1}^{c_2} \varphi(t)\, d\eta(t)$ es el límite de las «sumas integrales» (6.19).

3. Convergencia hacia el proceso estacionario. Examinemos el sistema lineal con la función de peso

$$w(t, s) = w(t - s), \quad t \geqslant s, \tag{6.26}$$

que sólo depende de la diferencia $(t - s)$ (la función $w(t)$, $t \geqslant 0$, también la llaman *función de peso*). Llamamos *estable* al sistema lineal si son cumplidas las siguientes condiciones:

$$\int_0^\infty |w(t)|\, dt < \infty, \quad \int_0^\infty |w(t)|^2\, dt < \infty. \tag{6.27}$$

Supongamos que sobre este sistema actúan las perturbaciones externas $\dot{\eta}(t)$ de tal tipo, que el proceso aleatorio $\xi(t)$ a la salida del sistema lo representamos por la integral estocástica

$$\xi(t) = \int_0^t w(t - s)\, d\eta(s), \tag{6.28}$$

donde el proceso aleatorio $\eta(t)$ con incrementos no correlacionados (que caracteriza la acción exterior integral, véase (6.8)) es *homogéneo* según el tiempo, en el sentido de que las funciones $a(t)$ y $b(t)$ de las relaciones (6.11) son constantes:

$$a(t) \equiv a, \quad b(t) \equiv b. \tag{6.29}$$

Examinemos el proceso aleatorio

$$\xi^0(t) = \int_{t_0}^t w(t - s)\, d\eta(s), \quad t_0 \leqslant t < \infty,$$

que empieza en el momento de tiempo t_0, y el proceso límite

$$\xi^*(t) = \int_{-\infty}^{t} w(t-s)\,d\eta(s) = \lim_{t_0 \to -\infty} \int_{t_0}^{t} w(t-s)\,d\eta(s), \quad -\infty < t < \infty \tag{6.30}$$

(que empezó en el pasado infinitamente lejano). Utilizando las fórmulas (6.23), (6.24), obtenemos

$$\| \xi(t) - \xi^*(t) \|^2 = \left\| \int_{-\infty}^{0} w(t-s)\,d\eta(s) \right\|^2 =$$

$$= b \int_{-\infty}^{0} |w(t-s)|^2\,ds + \left| a \int_{-\infty}^{0} w(t-s)\,ds \right|^2 =$$

$$= b \int_{t}^{\infty} |w(s)|^2\,ds + \left| a \int_{t}^{\infty} w(s)\,ds \right|^2;$$

se ve que

$$\| \xi(t) - \xi^*(t) \| \to 0 \quad \text{para} \quad t \to \infty. \tag{6.31}$$

Suponiendo que $w(t) = 0$ para $t < 0$, el proceso aleatorio $\xi^*(t)$ se puede representar en la forma

$$\xi^*(t) = \int_{-\infty}^{\infty} w(t-s)\,d\eta(s). \tag{6.32}$$

El valor medio de tal proceso es constante:

$$A(t) = \mathbf{M}\xi^*(t) = a \int_{-\infty}^{\infty} w(u)\,du; \tag{6.33}$$

la función de correlación

$$B(t, s) = \mathbf{M}\xi^*(t)\,\xi^*(s) = b \int_{-\infty}^{\infty} w(t-u)\,w(s-u)\,du =$$

$$= b \int_{-\infty}^{\infty} w(t-s-u)\,w(u)\,du = B(t-s) \tag{6.34}$$

depende sólo de la diferencia $t - s$. Tal proceso $\xi^*(t)$ se llama *estacionario (en un amplio sentido)*.

Nosotros hemos obtenido el siguiente resultado.

Teorema. *El proceso aleatorio* $\xi(t)$ *a la salida de un sistema estable, representado por la integral estocástica* (6.28), *cuando* $t \to \infty$, *converge en una media cuadrática hacia el proceso estacionario* $\xi^*(t)$ *determinado por la fórmula* (6.30).

Como complemento a esto señalamos además lo siguiente.

Supongamos que el proceso aleatorio $\eta(t)$ es un proceso *homogéneo con incrementos independientes*, es decir, las distribuciones de probabilidades de los incrementos $\eta(t) - \eta(s)$ son las mismas para todos los integrales (s, t) de igual longitud $t - s$, siendo las magnitudes $\eta(t_k) - \eta(s_k)$, $k = 1, \ldots, n$, mutuamente independientes, para cualesquiera $s_1 \leqslant t_1 \leqslant \ldots \leqslant s_n \leqslant t_n$.

Entonces, como se comprende fácilmente, el proceso $\xi^*(t) = \int_{-\infty}^{\infty} w(t-s)\, d\eta(s)$ será *estacionario* en el sentido de que la distribución conjunta de probabilidades de los valores $\xi(t_1 + s), \ldots, \xi(t_n + s)$ para cualesquiera valores fijados $t_1, \ldots, t_n$, es la misma para todos los valores de s. En particular, la distribución de probabilidades de los distintos valores $\xi^*(t)$, para todos los t, es la misma que para magnitud

$$\xi^* = \int_{-\infty}^{\infty} w(-u)\, d\eta(u). \tag{6.35}$$

De la convergencia de la media cuadrática $\| \xi(t) - \xi^*(t) \| \to 0$ se deduce (véase el teorema 2 del § 5, cap. I) que las distribuciones de probabilidades de las magnitudes $\xi(t)$ converge débilmente, para $t \to \infty$, hacia la distribución de probabilidades de la magnitud ξ^* determinada por la fórmula (6.35).

4. Procesos de efecto fraccionario. Examinemos el sistema lineal (6.1) al actuar sobre él los impulsos aleatorios $\eta_k \delta(t - \tau_k)$, $k = 1, 2, \ldots$, originados en los momentos aleatorios de tiempo τ_k, $k = 1, 2, \ldots$, Tal acción representa en sí al proceso aleatorio del tipo

$$\dot{\eta}(t) = \sum_{0 \leqslant \tau_k \leqslant t} \eta_k \delta(t - \tau_k), \quad 0 \leqslant t \leqslant \infty; \tag{6.36}$$

el estado del sistema $\xi(t)$ se cambiará durante el tiempo t de la forma siguiente:

$$\xi(t) = \sum_{0 \leqslant \tau_k \leqslant t} \eta_k w(t - \tau_k), \quad 0 \leqslant t \leqslant \infty \tag{6.37}$$

(compárese (6.4), (6.7)). El proceso aleatorio $\xi = \xi(t)$ de tal tipo se llama corrientemente *proceso de efecto fraccionario.*

Supongamos que los impulsos aleatorios $\eta_k \delta(t - \tau_k)$ surgen según la ley de Poisson con el parámetro λ, es decir, en el intervalo

de tiempo (s, t) con la probabilidad

$$\frac{[\lambda (t-s)]^n}{n!} e^{-\lambda (t-s)}$$

habrá justamente n impulsos (independientemente del número de impulsos fuera del intervalo dado desde s hasta t). Supongamos que las magnitudes aleatorias η_k, $k = 1, 2, \ldots$, que caracterizan la dirección y la intensidad de los impulsos correspondientes, tienen una misma distribución de probabilidades, siendo estas magnitudes tanto independientes entre sí, como de los momentos de tiempo τ_k, $k = 1, 2, \ldots$

La acción externa del tipo descrito se puede caracterizar, recurriendo al proceso aleatorio

$$\eta (t) = \int_0^t \dot{\eta} (s)\, ds = \sum_{k=1}^{\nu (t)} \eta_k, \quad 0 \leqslant t < \infty, \tag{6.38}$$

donde $\nu (t)$, $0 \leqslant t < \infty$, es el proceso corriente de Poisson y η_1, η_2, ... es la sucesión correspondiente de magnitudes aleatorias independientes con la misma distribución de probabilidades, también independientes de $\nu (t)$, $0 \leqslant t < \infty$. Tal proceso aleatorio $\eta = \eta (t)$ se denomina *proceso complejo de Poisson*.

Mostremos que con la condición de que las magnitudes correspondientes η_k tengan esperanza matemática y el segundo momento finito:

$$a = \mathbf{M}\eta_k, \quad b = \mathbf{M}\eta_k^2$$

se cumplirán las relaciones (6.11) siendo

$$a (t) = a\lambda t, \quad b (t) = b\lambda t, \tag{6.39}$$

(donde λ es el parámetro del proceso correspondiente de Poisson $\nu (t)$).

En efecto, si designamos por $\nu (s, t)$ el número de momentos τ_k en el intervalo $[s, t]$, entonces, para el valor fijado de $\nu (s, t) = n$, la magnitud $\eta = \eta (t) - \eta (s)$ reperesenta en sí la suma de n sumandos independientes η_k con un valor medio a y una dispersión $\sigma^2 = b - a^2$. Por consiguiente,

$$\mathbf{M} (\eta \mid \nu (s, t)) = a\nu (s, t),$$

y según la fórmula de la esperanza matemática total

$$\mathbf{M}\eta = \mathbf{M} \{\mathbf{M} (\eta \mid \nu (s, t))\} = a \cdot \mathbf{M}\nu (s, t) = a\lambda (t - s).$$

Luego,

$$\mathbf{M} (\eta^2 \mid \nu (s, t)) = \sigma^2 \nu (s, t) + (a\nu (s, t))^2,$$

$$\mathbf{M}\eta^2 = \mathbf{M} \{\mathbf{M} (\eta^2 \mid \nu (s, t))\} = \sigma^2 \lambda (t - s) + a^2 \mathbf{M} [\nu (s, t)]^2$$

y

$$\mathbf{D}\eta = \mathbf{M}\eta^2 - (\mathbf{M}\eta)^2 = \sigma^2\lambda(t-s) + a^2[\mathbf{M}\nu^2(s,\ t) - (\mathbf{M}\nu(s,\ t))^2] =$$
$$= \sigma^2\lambda(t-s) + a^2\mathbf{D}\nu(s,\ t) = \sigma^2\lambda(t-s) + a^2\lambda(t-s) =$$
$$= b\lambda(t-s).$$

Examinemos la integral estocástica $\int_{c_1}^{c_2} \varphi(t)\, d\eta(t)$, donde $\eta(t)$ es un proceso complejo de Poisson del tipo descrito anteriormente.

Mostremos, que para cualquier función continua a trozos $\varphi(t)$, $c_1 \leqslant t \leqslant c_2$, con la probabilidad 1

$$\int_{c_1}^{c_2} \varphi(t)\, d\eta(t) = \sum_{c_1 \leqslant \tau_j \leqslant c_2} \varphi(\tau_j)\, \eta_j. \tag{6.40}$$

(Recordemos que la propia integral estocástica $\xi = \int_{c_1}^{c_2} \varphi(t)\, d\eta(t)$ como límite medio cuadrático de una determinada sucesión de magnitudes aleatorias $\xi_n = \int_{c_1}^{c_2} \varphi_n(t)\, d\eta(t)$, $n = 1, 2, \ldots$, está solamente determinada con la probabilidad 1, a saber: cada magnitud que se diferencia de ξ con la probabilidad 0, también será límite de la misma sucesión ξ_n, $n = 1, 2, \ldots$).

Para la función constante a trozos $\varphi(t)$, $c_1 \leqslant t \leqslant c_2$, que conserva constantes los valores $y_k = \varphi(t)$ para $t_k < t < t_{k+1}$, donde los puntos $t_0 = c_1, \ldots, t_n = c_2$ dividen el segmento $[c_1, c_2]$ en un número finito de intervalos, la probabilidad de coincidencia de los momentos aleatorios τ_j con los puntos t_k, $k = 0, \ldots, n-1$, es igual a cero, y con la probabilidad 1

$$\sum_{c_1 \leqslant \tau_j \leqslant c_2} \varphi(\tau_j)\, \eta_j = \sum_{k=0}^{n-1} \Big[\sum_{t_k < \tau_j < t_{k+1}} \varphi(t_j)\, \eta_j \Big] =$$
$$= \sum_{k=0}^{n-1} \Big[y_k \sum_{t_k \leqslant \tau_j \leqslant t_{k+1}} \eta_j \Big] =$$
$$= \sum_{k=0}^{n-1} y_k [\eta(t_{k+1}) - \eta(t_k)] = \int_{c_1}^{c_2} \varphi(t)\, d\eta(t).$$

Para la función continua a trozos $\varphi(t)$, que es el límite de una sucesión u n i f o r m e m e n t e convergente de funciones continuas a trozos $\varphi_n(t)$, $n = 1, 2, \ldots$, la fórmula (6.40) se puede obte-

ner por un paso al límite:

$$\int_{c_1}^{c_2} \varphi_n(t)\, d\eta(t) = \sum_{c_1 \leqslant \tau_j \leqslant c_2} \varphi_n(\tau_j)\, \eta_j \to \int_{c_1}^{c_2} \varphi(t)\, d\eta(t),$$

y suponiendo, por otro lado, que

$$\Delta_n(t) = \varphi_n(t) - \varphi(t) \quad \text{y} \quad \varepsilon_n = \max_{c_1 \leqslant t \leqslant c_2} |\Delta_n(t)|,$$

tenemos

$$\Big\| \sum_{c_1 \leqslant \tau_j \leqslant c_2} \varphi_n(\tau_j)\, \eta_j - \sum_{c_1 \leqslant \tau_j \leqslant c_2} \varphi(\tau_j)\, \eta_j \Big\|^2 = \mathbf{M} \Big| \sum_{c_1 \leqslant \tau_j \leqslant c_2} \Delta_n(\tau_j)\, \eta_j \Big|^2 \leqslant \varepsilon_n^2 \mathbf{M}$$

$$\Big[\sum_{c_1 \leqslant \tau_j \leqslant c_2} |\eta_j| \Big]^2 \leqslant C\varepsilon_n^2 \to 0,$$

donde

$$C = \mathbf{M} \Big[\sum_{c_1 \leqslant \tau_j \leqslant c_2} |\eta_j| \Big]^2 = [\mathbf{M}|\eta_j|^2 + (\mathbf{M}|\eta_j|)^2]\, \lambda(c_2 - c_1).$$

La fórmula (6.40) que hemos obtenido nos permite representar el proceso de efecto fraccionario de la forma (6.37) por la integral estocástica del tipo (6.8):

$$\xi(t) = \int_0^t w(t, s)\, d\eta(s). \tag{6.41}$$

En particular, para la función de peso $w(t, s) = w(t - s)$, que satisface las condiciones (6.27), al proceso $\xi(t)$ de efecto fraccionario se le puede aplicar el teorema sobre la convergencia hacia el proceso estacionario (véase la pág. 209) y también su complemento, ya que el proceso complejo de Poisson $\eta(t)$, $t \geqslant 0$, en (6.41) es un proceso homogéneo con incrementos independientes.

¿Con qué probabilidad estará el sistema después del tiempo t, en tal o cual estado fásico? o más exactamente: ¿cuál será la distribución de probabilidades de la magnitud aleatoria $\xi(t)$?

Está claro, que esta distribución de probabilidades depende de la distribución de las magnitudes aleatorias η_k, $k = 1, 2, \ldots$, a la entrada del sistema. Sea $g(u)$ la función característica de las magnitudes η_k de igual distribución; hallemos la función característica $f_t(u) = \mathbf{M}e^{iu\xi(t)}$, $-\infty < u < \infty$, de la magnitud $\xi(t)$ que nos interesa.

Designemos por $\nu(t)$ el número de impulsos durante el tiempo t. Recordemos (véase el § 2, cap. II) que para el valor fijado $\nu(t) = n$, los momentos $\tau_1, \ldots, \tau_n$ de aparición de los impulsos están distribuidos en el segmento $[0, t]$ de tal modo como si los puntos $\tau_1, \ldots, \tau_n$ hubiesen sido lanzados casualmente, independientes uno del otro con la distribución uniforme de probabilidades de modo que para el

valor fijado $\nu(t) = n$, la magnitud $\xi(t)$ es la suma de n sumandos independientes $\eta_k w(t\tau_k)$, $k = 1, \ldots, n$ igualmente distribuidos (los τ_k están distribuidos uniformemente en el segmento $[0, t]$ y los η_k son independientes de $\nu(t)$ y $\tau_1, \tau_2, \ldots$). Por consiguiente, para el valor fijado $\tau_k = s$, el valor medio de la magnitud $e^{iu\,w(t,\tau_k)\,\eta_k}$ será

$$\mathbf{M}e^{iuw(t,\,s)\,\eta_k} = g(uw(t, s)),$$

y para el valor fijado $\nu(t) = n$

$$\mathbf{M}\{e^{iuw(t,\,\tau_k)\,\eta_k} \mid \nu(t) = n\} =$$

$$= \mathbf{M}\{g[uw(t, \tau_k)] \mid \nu(t) = n\} = \frac{1}{t}\int_0^t g[uw[t, s)]\,ds$$

y

$$\mathbf{M}\{e^{iu\xi(t)} \mid \nu(t) = n\} = \left[\frac{1}{t}\int_0^t g(uw(t, s))\,ds\right]^n.$$

Según la fórmula de la esperanza matemática total

$$\mathbf{M}e^{iu\xi(t)} = \sum_{n=0}^{\infty} \mathbf{M}\{e^{iu\xi(t)} \mid \nu(t) = n\}\,\mathbf{P}\{\nu(t) = n\} =$$

$$= \sum_{n=0}^{\infty}\left[\frac{1}{t}\int_0^t g(uw(t, s))\,ds\right]^n \frac{(\lambda t)^n}{n!}e^{-\lambda t} = e^{-\lambda t}\left[e^{\lambda\int_0^t g(uw(t,\,s))\,ds}\right].$$

Como resultado, *para la función característica* $f_t(u) = \mathbf{M}e^{iu\,\xi(t)}$ *se tiene la siguiente fórmula*:

$$f_t(u) = \exp\left\{\lambda\int_0^t [g(uw(t, s)) - 1]\,ds\right\}, \quad -\infty < u < \infty, \quad (6.42)$$

donde $g(u) = \mathbf{M}e^{iu\eta_k}$ *es la función característica de las magnitudes* η_k, $k = 1, 2, \ldots$, y λ *es el parámetro de la ley de Poisson correspondiente.*

En particular, para $w(t) \equiv 1$, $0 \leqslant t < \infty$, cuando tratamos con *el proceso complejo de Poisson* $\xi(t) = \eta(t)$ (véase (6.38)), la función característica correspondiente $f_t(u)$ será

$$f_t(u) = e^{\lambda t\,[g(u)-1]}, \quad -\infty < u < \infty \quad (6.43)$$

(para el proceso corriente de Poisson $\eta_k \equiv 1$ y $g(u) = e^{iu}$).

En resumen señalemos, que para el sistema lineal estable con la función de peso $w(t)$, cuando el proceso aleatorio $\xi(t)$ converge hacia el proceso estacionario para $t \to \infty$, (véase (6.31)), la distribución

de probabilidades de las magnitudes $\xi(t)$ converge débilmente hacia la distribución de la magnitud ξ^*, determinada por la fórmula (6.35), y por consiguiente, las funciones características de las magnitudes $\xi(t)$, que según (6.42) tienen la forma

$$f_t(u) = \exp\left\{\lambda \int_0^t [g(uw(s)) - 1]\, ds\right\}, \quad -\infty < u < \infty,$$

convergen hacia la función característica de la magnitud ξ^* para $t \to \infty$

$$f(u) = \exp\left\{\lambda \int_0^\infty [g(uw(s)) - 1]\, ds\right\}, \quad -\infty < u < \infty.$$

§ 7. PROCESOS ALEATORIOS ESTACIONARIOS

1. Representación espectral de los procesos estacionarios y transformación de Fourier. Recordemos que el proceso aleatorio $\xi = \xi(t)$, $-\infty < t < \infty$, se llama estacionario, en amplio sentido, si el valor medio $A(t) = \mathbf{M}\xi(t)$ es uno mismo para todos los valores de t, y la función de correlación $B(t, s) = \mathbf{M}\xi(t)\,\overline{\xi(s)}$, sólo depende de la diferencia $t - s$. En adelante consideraremos que $A(t) = 0$.

Ampliemos el concepto de estacionaridad, extendiéndolo a los procesos aleatorios con valores complejos $\xi(t)$, $\mathbf{M}\xi(t) = 0$ del modo siguiente: el proceso aleatorio $\xi(t)$, $-\infty < t < \infty$, se llama *estacionario en amplio sentido*, si $\mathbf{M}\xi(t)\,\overline{\xi(s)}$ sólo depende de la diferencia $t - s$

$$\mathbf{M}\xi(t)\,\overline{\xi(s)} = B(t - s); \tag{7.1}$$

la función $B(t)$, $-\infty < t < \infty$, se llama *función de correlación* del proceso estacionario $\xi(t)$.

Ejemplo (proceso estacionario con espectro discreto). Examinemos el proceso aleatorio de la forma

$$\xi(t) = \sum_{k=-n}^{n} e^{i\lambda_k t}\, \Phi_k, \quad -\infty < t < \infty \tag{7.2}$$

sonde $\lambda_{-n}, \ldots, \lambda_n$ son determinados números reales, y $\Phi_{-n}, \ldots, \Phi_n$ don magnitudes aleatorias no correlacionadas con valores medios nulos:

$$\mathbf{M}\Phi_k\overline{\Phi}_j = 0 \quad \text{para} \quad k \neq j.$$

Evidentemente,

$$\mathbf{M}\xi(t) = \sum_{k=-n}^{n} e^{i\lambda_k t}\,\mathbf{M}\Phi_k = 0$$

y

$$\mathbf{M}\xi(t)\,\overline{\xi(s)} = \sum_{k=-n}^{n}\sum_{j=-n}^{n} e^{i(\lambda_k t - \lambda_j s)}\,\mathbf{M}\Phi_k\overline{\Phi}_j = \sum_{k=-n}^{n} e^{i\lambda_k (t-s)}\,F_k, \quad (7.3)$$

donde $F_k = \mathbf{M}\,|\,\Phi_k\,|^2$, $k = -n, \ldots, n$, y de tal modo, el proceso aleatorio $\xi(t)$ es estacionario.

Este proceso da un ejemplo de oscilaciones aleatorias que tienen como componentes suyas a las oscilaciones armónicas $\xi_k(t) = e^{i\lambda_k t}\,\Phi_k$ con las frecuencias $\omega_k = |\,\lambda_k\,|$, $k = -n, \ldots, n$, cuya face y amplitud son magnitudes aleatorias. Si convenimos en llamar al valor medio cuadrático $\mathbf{M}\,|\,\xi(t)\,|^2$ la energía media del proceso estacionario $\xi(t)$, entonces de la igualdad (7.3), para $s = t$, tenemos

$$\mathbf{M}\,|\,\xi(t)\,|^2 = \sum_{k=-n}^{n} F_k,$$

donde $F_k = \mathbf{M}\,|\,\Phi_k\,|^2 = \mathbf{M}\,|\,\xi_k(t)\,|^2$, y se ve que la energía del proceso estacionario $\xi(t)$ se compone de las energías de las componentes separadas $\xi_k(t) = e^{i\lambda_k t}\,\Phi_k$, $k = -n, \ldots, n$.

Los valores λ_k, $k = -n, \ldots, n$, se llaman corrientemente, *frecuencias* (incluso para los valores negativos de λ_k); en su conjunto forman *el espectro de frecuencias* del proceso estacionario $\xi(t)$.

Ejemplo (*proceso estacionario con espectro continuo*). Examinemos el proceso aleatorio $\xi(t)$ que admite la llamada *representación espectral*:

$$\xi(t) = \int_{-\infty}^{\infty} e^{\lambda t}\,d\Phi(\lambda), \quad (7.4)$$

donde $\Phi(\lambda)$, $-\infty < \lambda < \infty$ es el proceso aleatorio con incrementos no correlacionados y con un valor medio nulo tal que para cualquier intervalo (λ_1, λ_2)

$$\mathbf{M}\,|\,\Phi(\lambda_1) - \Phi(\lambda_2)\,|^2 = \int_{\lambda_1}^{\lambda_2} f(\lambda)\,d\lambda; \quad (7.5)$$

la densidad de estructura $f(\lambda)$ de este proceso se supone que es integrable.

En la fórmula (7.4) figura la integral estocástica de la función compleja $\varphi(\lambda) = e^{i\lambda t}$, $-\infty < \lambda < \infty$.

Recordemos, que la integral estocástica $\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda)$ definida por nosotros en el § 6, posee las siguientes propiedades:

$$\mathbf{M} \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda) = 0,$$

$$\mathbf{M} \left| \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda) \right|^2 = \int_{-\infty}^{\infty} |\varphi(\lambda)|^2 f(\lambda)\, d\lambda, \tag{7.6}$$

y en general,

$$\mathbf{M} \left[\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda) \right] \overline{\left[\int_{-\infty}^{\infty} \psi(\lambda)\, d\Phi(\lambda) \right]} = \int_{-\infty}^{\infty} \varphi(\lambda) \overline{\psi(\lambda)} f(\lambda)\, d\lambda, \tag{7.7}$$

donde la función $f(\lambda) \geqslant 0$ es la densidad de estructura del proceso $\Phi(\lambda)$ que figura en la relación (7.5).

Así pues, examinemos el proceso aleatorio $\xi(t)$ determinado por la igualdad (7.4). Para $\varphi(\lambda) = e^{i\lambda t}$, $\psi(\lambda) = e^{i\lambda s}$ obtenemos de las fórmulas (7.6) y (7.7) que

$$\mathbf{M}\xi(t) = 0$$

y

$$\mathbf{M}\xi(t)\overline{\xi(s)} = \int_{-\infty}^{\infty} e^{i\lambda(t-s)} f(\lambda)\, d\lambda.$$

Se ve que $\xi(t)$ es un proceso estacionario con la función de correlación

$$B(t) = \int_{-\infty}^{\infty} e^{i\lambda t} f(\lambda)\, d\lambda, \tag{7.8}$$

que coincide con la transformación de Fourier de la densidad de estructura $f(\lambda)$ del proceso $\Phi(\lambda)$; $f(\lambda)$ también se llama *densidad espectral* del proceso estacionario $\xi(t)$. A su vez, la densidad espectral $f(\lambda)$ se determina unívocamente, por su transformación de Fourier $B(t)$, y, en particular, cuando la función de correlación $B(t)$, $-\infty < t < \infty$, es integrable

$$f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} B(t)\, dt. \tag{7.9}$$

La fórmula (7.4) da la descomposición del proceso estacionario $\xi(t)$ en oscilaciones armónicas elementales, más exacto,

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda) \sim \sum_{k=-n}^{n} e^{i\lambda_k t}\, \Phi_k,$$

donde $\Phi_k = \Phi(\lambda_{k+1}) - \Phi(\lambda_k)$, $k = -n, \ldots, n$, y el signo de equivalencia $\sim$ significa, que para la división cada vez más pequeña, de las bandas de frecuencias $-\Lambda_n \leqslant \lambda \leqslant \Lambda_n$ en intervalos $\Delta_k = (\lambda_k, \lambda_{k+1})$ $k = -n, \ldots, n$,

$$\xi(t) = \text{l.i.m.}_{n\to\infty} \sum_{k=-n}^{n} e^{i\lambda_k t}\, \Phi_k \tag{7.10}$$

(para $\max\limits_k |\lambda_{k+1} - \lambda_k| \to 0$, $\Lambda_n \to \infty$).

Señalemos que para cualquier intervalo de frecuencias $\lambda_1 \leqslant \lambda \leqslant \lambda_2$, el proceso aleatorio estacionario

$$\xi_\Delta(t) = \int_{\lambda_1}^{\lambda_2} e^{i\lambda t}\, d\Phi(\lambda), \tag{7.11}$$

en un pequeño intervalo $\Delta = (\lambda_1, \lambda_2)$ representa en sí aproximadamente, una oscilación armónica de frecuencia λ, $\lambda_1 \leqslant \lambda \leqslant \lambda_2$ (tanto más exacto en el intervalo de tiempo fijado $t_0 \leqslant t \leqslant t_1$, cuanto menor sea Δ) y su energía media es

$$\mathbf{M}\,|\xi_\Delta(t)|^2 = \int_{\lambda_1}^{\lambda_2} f(\lambda)\, d\lambda. \tag{7.12}$$

La energía total del proceso estacionario $\xi(t)$ es

$$\mathbf{M}\,|\xi(t)|^2 = \int_{-\infty}^{\infty} f(\lambda)\, d\lambda.$$

y de este modo, como indica la fórmula (7.12), la densidad espectral $f(\lambda)$ caracteriza la distribución de la energía del proceso examinado $\xi(t) = \int\limits_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda)$, según las componentes de la forma $\xi_\Delta(t) = \int\limits_{\lambda_1}^{\lambda_2} e^{i\lambda t} d\Phi(\lambda)$ en dependencia del intervalo $\Delta = (\lambda_1, \lambda_2)$ de aquellas frecuencias λ, que entran en el espectro de frecuencias de las oscilaciones de la componente correspondiente $\xi_\Delta(t)$.

Examinemos el proceso aleatorio estacionario de la forma

$$\xi(t) = \int_{\infty}^{t} w(t-s)\, d\eta(s), \tag{7.13}$$

que se establece en el transcurso del tiempo a la salida del sistema lineal estable con la función de peso $w(t)$, al actuar sobre ella las perturbaciones aleatorias homogéneas $\dot{\eta}(t)$ (véase el teorema del § 6 anterior).

Recordemos que hemos llegado al proceso estacionario del tipo (7.13) con ayuda del paso límite:

$$\xi(t) = \text{l.i.m.}_{t_0 \to -\infty} \int_{t_0}^{t} w(t-s)\, d\eta(s),$$

desplazando el momento inicial t_0 del funcionamiento del sistema examinado, al pasado infinitamente lejano. Para cualesquiera valores de $t_0 \leqslant s \leqslant t$, los incrementos

$$\eta(t) - \eta(s) = \int_{s}^{t} \dot{\eta}(u)\, du,$$

que caracterizan la acción total exterior sobre el sistema en el intervalo de tiempo desde s hasta t, no depende de la elección del origen de lectura t_0. A pesar de que los propios valores $\eta(t)$ pierden su sentido, para $t_0 \to -\infty$ de hecho tratamos sólo con las magnitudes $\Delta\eta = \eta(t) - \eta(s)$, determinadas para cualquier intervalo $\Delta = (s, t]$, que poseen la misma propiedad que para la división del intervalo $\Delta = (s, t]$ en intervalos que no se cortan $\Delta_k = (s_k, t_k]$, $k = 1, \ldots, n$ ($\Delta = \bigcup_{k=1}^{n} \Delta_k$), tiene lugar la igualdad

$$\Delta\eta = \sum_{k=1}^{n} \Delta_k \eta.$$

Para mayor claridad, hablaremos sobre el proceso aleatorio homogéneo $\eta(t)$ con incrementos no correlacionados $\Delta\eta = \eta(t) - \eta(s)$, y en el caso $t_0 = -\infty$, suponiendo que

$$\left.\begin{aligned} \mathbf{M}\Delta\eta = 0, \quad \mathbf{M}|\Delta\eta|^2 = \sigma^2(t-s), \\ \mathbf{M}(\Delta_1\eta \cdot \overline{\Delta_2\eta}) = 0 \end{aligned}\right\} \tag{7.14}$$

para cualesquiera intervalos Δ_1, Δ_2 que no se cortan; la homogeneidad del proceso $\eta(t)$, mencionada anteriormente, significa que su densidad de estructura es igual a una constante (en nuestro caso esta constante se ha designado por σ^2).

Examinemos ahora el proceso aleatorio $\psi(\lambda)$ determinado para todos los $\lambda \geqslant \lambda_0$ como

$$\psi(\lambda_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\lambda t} - e^{-i\lambda_0 t}}{-it}\, d\eta(t).$$

Para cualesquiera $\lambda_0 \leqslant \lambda_1 \leqslant \lambda_2$ los incrementos $\Delta\psi = \psi(\lambda_2) - \psi(\lambda_1)$ en el intervalo $\Delta = (\lambda_1, \lambda_2)$ representados en la forma

$$\Delta\psi = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\lambda_2 t} - e^{-i\lambda_1 t}}{-it}\, d\eta(t). \tag{7.15}$$

poseen las siguientes propiedades:

$$\left.\begin{aligned} \mathbf{M}\Delta\psi = 0, \quad \mathbf{M}|\Delta\psi|^2 = \frac{\sigma^2}{2\pi}(\lambda_2 - \lambda_2). \\ \mathbf{M}(\Delta_1\psi \cdot \overline{\Delta_2\psi}) = 0 \end{aligned}\right\} \tag{7.16}$$

para los intervalos Δ_1, Δ_2 que no se cortan. Las dos últimas relaciones que necesitan su demostración se obtienen fácilmente al utilizar la igualdad de Parceval; precisamente, la función subintegral

$$e_\Delta(t) = \frac{1}{2\pi} \frac{e^{-i\lambda_2 t} - e^{-i\lambda_1 t}}{-it}, \qquad \int_{-\infty}^{\infty} |e_\Delta(t)|^2\, dt < \infty,$$

es la transformación de Fourier de la forma

$$e_\Delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t}\, \chi_\Delta(\lambda)\, d\lambda, \tag{7.17}$$

donde $\chi_\Delta(\lambda)$ es el indicador del intervalo $\Delta = (\lambda_1, \lambda_2]$ ($\chi_\Delta(\lambda) = 1$ para $\lambda \in \Delta$ y $\chi_\Delta(\lambda) = 0$ para $\lambda \notin \Delta$) y

$$\mathbf{M}|\Delta\psi|^2 = \sigma^2 \int_{-\infty}^{\infty} |e_\Delta(t)|^2\, dt = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} |\chi_\Delta(\lambda)|^2\, d\lambda = \frac{\sigma^2}{2\pi}(\lambda_2 - \lambda_1).$$

$$\mathbf{M}(\Delta_1\psi \cdot \overline{\Delta_2\psi}) = \sigma^2 \int_{-\infty}^{\infty} e_{\Delta_1}(t)\, \overline{e_{\Delta_2}(t)}\, dt = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} \chi_{\Delta_1}(\lambda)\, \overline{\chi_{\Delta_2}(\lambda)}\, d\lambda = 0$$

para los cualesquiera intervalos Δ_1, Δ_2 que no se cortan.

La fórmula (7.15) indica que los incrementos $\Delta\psi = \psi(\lambda_2) - \psi(\lambda_1)$ son los mismos para cualquier valor inicial λ_0. Por eso, para las funciones $\varphi(\lambda)$ que satisfagan a la condición

$$\int_{-\infty}^{\infty} |\varphi(\lambda)|^2\, d\lambda < \infty,$$

está determinada la integral estocástica

$$\int_{-\infty}^{\infty} \varphi(\lambda)\, d\psi(\lambda) = \text{l.i.m.}_{\lambda_0 \to -\infty} \int_{\lambda_0}^{\infty} \varphi(\lambda)\, d\psi(\lambda).$$

En el mismo sentido que sobre el proceso $\eta(t)$, $-\infty < t < \infty$, hablaremos del proceso aleatorio $\psi(\lambda)$, $-\infty < \lambda < \infty$, con incrementos no correlacionados, que tiene un valor medio nulo y una densidad de estructura constante $g(\lambda) = \sigma^2/2\pi$, aunque tienen sólo sentido real los «incrementos» $\Delta\psi = \psi(\lambda_2) - \psi(\lambda_1)$ determinados por la fórmula (7.15) para cualquier intervalo $\Delta = (\lambda_1, \lambda_2]$ (a este proceso aleatorio generalizado $\psi(\lambda)$ lo llaman corrientemente, *transformación de Fourier* del proceso aleatorio $\eta(t)$.

Consideremos la integral estocástica $\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Psi(\lambda)$.

Para la función constante a trozos de la forma $\varphi(\lambda) = \sum_{k=1}^{n} y_k \chi_{\Delta_k}(\lambda)$, donde $\Delta_k = (\lambda_k', \lambda_k'')$, $k = 1, \ldots, n$ son los intervalos que no se cortan de longitud finita y su transformación de Fourier

$$c(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda t} \varphi(\lambda)\, d\lambda \tag{7.18}$$

tenemos

$$\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Psi(\lambda) = \int_{-\infty}^{\infty} c(t)\, d\eta(t), \tag{7.19}$$

lo que es un corolario simple de las fórmulas (7.15), (7.17), a saber:

$$\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Psi(\lambda) = \sum_{k=1}^{n} y_k \Delta_k \Psi = \int_{-\infty}^{\infty} \left[\sum_{k=1}^{n} y_k e_{\Delta_k}(t) \right] d\eta(t) = \int_{-\infty}^{\infty} c(t)\, d\eta(t).$$

La distancia media cuadrática entre las integrales estocásticas del tipo $\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Psi(\lambda)$ es

$$\mathbf{M} \left| \int_{-\infty}^{\infty} \varphi_1(\lambda)\, d\Psi(\lambda) - \int_{-\infty}^{\infty} \varphi_2(\lambda)\, d\Psi(\lambda) \right|^2 = \frac{\sigma^2}{2\pi} \int_{-\infty}^{\infty} |\varphi_1(\lambda) - \varphi_2(\lambda)|^2\, d\lambda.$$

Análogamente

$$\mathbf{M} \left| \int_{-\infty}^{\infty} c_1(t)\, d\eta(t) - \int_{-\infty}^{\infty} c_2(t)\, d\eta(t) \right|^2 = \sigma^2 \int_{-\infty}^{\infty} |c_1(t) - c_2(t)|^2\, dt.$$

Recordemos que para las funciones $c(t)$, $\int\limits_{-\infty}^{\infty} |c(t)|^2\,dt < \infty$, que son transformaciones de Fourier de la forma (7.18) de las funciones $\varphi(\lambda)$ integrables absolutamente, según la igualdad de Parceval

$$\int\limits_{-\infty}^{\infty} |c(t)|^2\,dt = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |\varphi(\lambda)|^2\,d\lambda;$$

por eso

$$\frac{1}{2\pi}\int\limits_{-\infty}^{\infty} |\varphi_1(\lambda)-\varphi_2(\lambda)|^2\,d\lambda = \int\limits_{-\infty}^{\infty} |c_1(t)-c_2(t)|^2\,dt$$

y

$$\mathbf{M}\left|\int\limits_{-\infty}^{\infty}\varphi_1(\lambda)\,d\Psi(\lambda)-\int\limits_{-\infty}^{\infty}\varphi_2(\lambda)\,d\Psi(\lambda)\right|^2 = \mathbf{M}\left|\int\limits_{-\infty}^{\infty} c_1(t)\,d\eta(t)-\int\limits_{-\infty}^{\infty} c_2(t)\,d\eta(t)\right|^2,$$

donde $c_1(t)$ y $c_2(t)$ son transformaciones de Fourier del tipo (7.18) de las funciones $\varphi_1(\lambda)$ y $\varphi_2(\lambda)$.

Está claro, que si la función $\varphi(\lambda)$ es el límite medio cuadrático de la sucesión de funciones constantes a trozos $\varphi_n(\lambda)$, $n = 1, 2, \ldots$, entonces su transformación de Fourier $c(t)$ coincide con el límite medio cuadrático de la sucesión correspondiente $c_n(t)$, $n = 1, 2, \ldots$, siendo

$$\int\limits_{-\infty}^{\infty}\varphi(\lambda)\,d\Psi(\lambda) = \text{l.i.m.}_{n\to\infty}\int\limits_{-\infty}^{\infty}\varphi_n(\lambda)\,d\Psi(\lambda),$$

$$\int\limits_{-\infty}^{\infty} c(t)\,d\eta(t) = \text{l.i.m.}_{n\to\infty}\int\limits_{-\infty}^{\infty} c_n(t)\,d\eta(t).$$

Sin embargo, para las funciones constantes a trozos $\varphi_n(\lambda)$ tiene lugar la igualdad (7.19)

$$\int\limits_{-\infty}^{\infty}\varphi_n(\lambda)\,d\Psi(\lambda) = \int\limits_{-\infty}^{\infty} c_n(t)\,d\eta(t), \quad n = 1, 2, \ldots,$$

y, por consiguiente,

$$\int\limits_{-\infty}^{\infty}\varphi(\lambda)\,d\Psi(\lambda) = \text{l.i.m.}_{n\to\infty}\int\limits_{-\infty}^{\infty}\varphi_n(\lambda)\,d\Psi(\lambda) = \text{l.i.m.}_{n\to\infty}\int\limits_{-\infty}^{\infty} c_n(t)\,d\eta(t) = \int\limits_{-\infty}^{\infty} c(t)\,d\eta(t).$$

De tal modo, la igualdad (7.19) tiene lugar para cualesquiera pares $\varphi(\lambda)$ y $c(t)$ que satisfagan las condiciones

$$\int_{-\infty}^{\infty} |\varphi(\lambda)|^2 d\lambda < \infty, \quad \int_{-\infty}^{\infty} |c(t)|^2 dt < \infty$$

(donde $c(t)$ es la transformación de Fourier de la forma (7.18) de la función $\varphi(\lambda)$).

Partiendo de aquellas funciones $\varphi(\lambda)$ y $c(t)$ que satisfacen las condiciones complementarias

$$\int_{-\infty}^{\infty} |\varphi(\lambda)|^2 d\lambda < \infty, \quad \int_{-\infty}^{\infty} |c(t)| dt < \infty \tag{7.20}$$

(en este caso la transformación inversa de Fourier nos da $\varphi(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} c(t)\, dt$), con ayuda del paso límite, la igualdad (7.19) se puede extender a *cualesquiera funciones* $c(t)$, *absolutamente integrables* y a sus transformaciones de Fourier

$$\varphi(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} c(t)\, dt.$$

Consideremos ahora directamente al proceso estacionario $\xi(t)$, representado por la integral estocástica (7.13):

$$\xi(t) = \int_{-\infty}^{\infty} w(t-s)\, d\eta(s).$$

Llamemos *característica espectral*[1]) del sistema lineal con la función de peso $w(t)$, ($w(t)$, $-\infty < t < \infty$ es absolutamente inte-

---

[1]) En las aplicaciones el sistema lineal se caracteriza, corrientemente, por la tal llamada *función de transmisión* $\varphi(p)$ que está ligada a la función de peso $w(t)$ correspondiente ($w(t) = 0$ para $t < 0$), con la transformación de Laplace:

$$\varphi(p) = \int_0^{\infty} e^{-pt} w(t)\, dt$$

(*véase, por ejemplo, el libro de* J. Lanning y R. Battin «Procesos aleatorios en los sistemas de dirección automática» (trad. al ruso del inglés), IL, Moscú, 1958.

grable), a la función

$$\varphi(i\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda t} w(t)\, dt. \tag{7.21}$$

Supongamos que la característica espectral $\varphi(i\lambda)$ es tal que $\int_{-\infty}^{\infty} |\varphi(i\lambda)|^2\, d\lambda < \infty$. Evidentemente, la función $c(s) = w(t-s)$, $-\infty < s < \infty$, satisface a las condiciones (7.20) siendo

$$\int_{-\infty}^{\infty} e^{-i\lambda s} c(s)\, ds = e^{i\lambda t} \int_{-\infty}^{\infty} e^{-i\lambda(t-s)} w(t-s)\, ds = e^{i\lambda t}\varphi(i\lambda).$$

Por consiguiente, de acuerdo con la fórmula general (7.19), tenemos

$$\xi(t) = \int_{-\infty}^{\infty} w(t-s)\, d\eta(s) = \int_{-\infty}^{\infty} e^{i\lambda t}\varphi(i\lambda)\, d\Psi(\lambda). \tag{7.22}$$

La igualdad

$$\Phi(\lambda) = \int_{-\infty}^{\lambda} \varphi(i\lambda)\, d\Psi(\lambda) \tag{7.23}$$

determina el proceso aleatorio complejo con incrementos no correlacionados, que tiene el valor medio nulo y la densidad de estructura

$$f(\lambda) = \frac{\sigma^2}{2\pi} |\varphi(i\lambda)|^2. \tag{7.24}$$

En efecto, designando por $\Delta\Phi = \Phi(\lambda_2) - \Phi(\lambda_1)$ el incremento en el intervalo correspondiente $\Delta = (\lambda_1, \lambda_2)$ tenemos

$$\mathbf{M}\Delta\Phi = 0,$$

$$\mathbf{M}|\Delta\Phi|^2 = \mathbf{M}\left|\int_{-\infty}^{\infty} \varphi(i\lambda)\chi_\Delta(\lambda)\, d\Psi(\lambda)\right|^2 = \frac{\sigma^2}{2\pi}\int_{\lambda_1}^{\lambda_2} |\varphi(i\lambda)|^2\, d\lambda,$$

$$\mathbf{M}(\Delta_1\Phi\overline{\Delta_2\Phi}) = \mathbf{M}\left[\int_{-\infty}^{\infty} \varphi(i\lambda)\varkappa_{\Delta_1}(\lambda)\, d\Psi(\lambda) \overline{\int_{-\infty}^{\infty} \varphi(i\lambda)\varkappa_{\Delta_2}\, d\Psi(\lambda)}\right] =$$

$$= \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty} |\varphi(i\lambda)|^2 \chi_{\Delta_1}(\lambda)\overline{\chi_{\Delta_2}(\lambda)}\, d\lambda = 0$$

para cualesquiera intervalos que no se cortan Δ_1, Δ_2.

De la definición general de integral estocástica se deduce fácilmente que

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\varphi(i\lambda)\, d\Psi(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda). \tag{7.25}$$

Hemos obtenido la representación espectral del proceso estacionario $\xi(t)$ (véase (7.4) y más adelante).

Así pues, tiene lugar el siguiente resultado.

Teorema. *Cualquier proceso estacionario aleatorio* $\xi(t)$ *del tipo* (7.13) *admite la representación espectral* (7.4); *con ello su densidad espectral* $f(\lambda)$ *se expresa por la fórmula* (7.24).

2. **Transformaciones lineales. Ejemplos.** Sea $\xi(t)$, $-\infty < t < \infty$, un proceso estacionario aleatorio, que admite la representación espectral

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda).$$

Sea $T = [a, b]$ un segmento finito o infinito sobre el eje del tiempo $-\infty < t < \infty$. Designemos por $L_T(f)$ al conjunto de todas las funciones $\varphi(\lambda)$, $-\infty < \lambda < \infty$, que tienen bien la forma $\sum_{k=1}^{n} c_k e^{i\lambda t_k}$ (donde $t_1, \ldots, t_n \in T$), o bien son límites para tales funciones:

$$\varphi(\lambda) = \text{l.i.m.} \sum_k c_k e^{i\lambda t_k} \tag{7.26}$$

en el sentido de *la convergencia media cuadrática* con relación a la distancia

$$\| \varphi_1(\lambda) - \varphi_2(\lambda) \| = \left(\int_{-\infty}^{\infty} |\varphi_1(\lambda) - \varphi_2(\lambda)|^2 f(\lambda)\, d\lambda \right)^{1/2},$$

donde $f(\lambda)$ es la densidad espectral del proceso estacionario examinado $\xi(t)$.

Designemos por $H(T)$ al conjunto de todas las magnitudes de la forma

$$\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda), \tag{7.27}$$

donde $\varphi(\lambda) \in L_T(f)$. Para $\varphi(\lambda) = \sum_k c_k e^{i\lambda t_k}$ las magnitudes indicadas representan en sí las combinaciones lineales de los valores correspon-

dientes $\xi(t_k)$, $t_k \in T$:

$$\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda) = \int_{-\infty}^{\infty} \Big[\sum_k c_k e^{i\lambda t_k}\Big] d\Phi(\lambda) =$$

$$= \sum_k c_k \int_{-\infty}^{\infty} e^{i\lambda t_k}\, d\Phi(\lambda) = \sum_k c_k \xi(t_k)$$

y en caso general son límites (con relación a la convergencia media cuadrática) para tales combinaciones lineales:

$$\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda) = \text{l.i.m.} \sum_k c_k \xi(t_k). \tag{7.28}$$

Ejemplo (*integración*). Sea $c(t)$, $-\infty < t < \infty$, una función continua a trozos, igual a cero fuera del segmento finito $T = [a, b]$ y sea $\varphi(\lambda)$ su transformación de Fourier:

$$\varphi(\lambda) = \int_a^b e^{i\lambda t} c(t)\, dt = \lim \sum_{k=0}^{n-1} e^{i\lambda t_k} c(t_k) \Delta t_k.$$

donde $\Delta t_k = t_{k+1} - t_k$, $k = 0, 1, \ldots, n-1$, y el límite «de las sumas integrales» $\varphi_n(\lambda) = \sum_{k=0}^{n-1} e^{i\lambda t_k} c(t_k)\, \Delta t_k$ se toma, como corrientemente, para la división cada vez más pequeña del segmento $[a, b]$ por los puntos $t_0 = a, t_1, \ldots, t_n = b$ (cuando $\max_k \Delta t_k \to 0$). Se ve fácilmente, que la relación límite indicada también se cumple en el sentido de la convergencia media cuadrática:

$$\text{l.i.m.} \int_{-\infty}^{\infty} |\varphi(\lambda) - \varphi_n(\lambda)|^2 f(\lambda)\, d\lambda = 0. \tag{7.29}$$

En efecto, para cualquiera $\varepsilon > 0$ se puede elegir tal a que $\int_{|\lambda|>a} f(\lambda)\, d\lambda \leqslant \varepsilon$ (recordemos, que la densidad espectral $f(\lambda)$ es una función integrable); la sucesión $\varphi_n(\lambda) = \sum_{k=1}^{n} e^{i\lambda t_k} c(t_k)\, \Delta t_k$ está limitada y converge hacia $\varphi(\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} c(t)\, dt$ uniformemente por λ en cada intervalo finito:

$$|\varphi(\lambda) - \varphi_n(\lambda)| \leqslant \varepsilon \quad \text{para} \quad |\lambda| \leqslant a;$$

para todos los valores de n suficientemente grandes (digamos, para $n \geqslant n_\varepsilon$) tenemos

$$\int_{-\infty}^{\infty} |\varphi(\lambda) - \varphi_n(\lambda)|^2 f(\lambda)\, d\lambda = \int_{|\lambda|>a} |\varphi(\lambda) - \varphi_n(\lambda)|^2 f(\lambda) +$$

$$+ \int_{-a}^{a} |\varphi(\lambda) - \varphi_n(\lambda)|^2 f(\lambda)\, d\lambda \leqslant C \int_{|\lambda|>a} f(\lambda)\, d\lambda + \varepsilon \int_{-a}^{a} f(\lambda)\, d\lambda \leqslant C_1 \varepsilon,$$

cuando $n \geqslant n_\varepsilon$ para cualquier valor de $\varepsilon > 0$ dado de antemano (donde C, C_1 son determinados constantes).

Examinemos la magnitud correspondiente $\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda)$ del espacio $H(T)$:

$$\eta = \text{l.i.m.} \int_{-\infty}^{\infty} \varphi_n(\lambda)\, d\Phi(\lambda) = \text{l.i.m.} \sum_{k=1}^{n} c(t_k)\, \xi(t_k)\, \Delta_n t_k.$$

Llamemos *integral* al límite medio cuadrático de "las sumas integrales" $\sum_{k=1}^{n} c(t_k)\, \xi(t_k)\, \Delta t_k$ suponiendo [1]) que

$$\int_{-\infty}^{\infty} c(t)\, \xi(t)\, dt = \text{l.i.m.} \sum_{k=1}^{n} c(t_k)\, \xi(t_k)\, \Delta t_k.$$

En estas designaciones

$$\int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda) = \int_{-\infty}^{\infty} c(t)\, \xi(t)\, dt. \tag{7.30}$$

Por un paso límite la fórmula (7.30) se extiende a las funciones arbitrarias absolutamente integrables $c(t)$, $-\infty < t < \infty$, y a sus transformaciones de Fourier $\varphi(\lambda)$, $-\infty < \lambda < \infty$ (véase (7.19)).

[1]) Señalemos, que para el proceso aleatorio $\eta(t)$, $a \leqslant t \leqslant b$, en el cual todas trayectorias posibles son absolutamente integrables (para nosotros $\eta(t) = c(t)\, \xi(t)$, el límite medio cuadrático de «las sumas integrales» $\text{l.i.m.} \sum_{k=1}^{n} \eta(t_k)\, \Delta t_k = \int_a^b \eta(t)\, dt$ coincide, como magnitud aleatoria con probabilidad 1, con la integral corriente $\int_a^b \eta(t)\, dt$ de la trayectoria existente $\eta(t)$, $a \leqslant t \leqslant b$ (véase, por ejemplo el cap. IV del libro citado anteriormente, de I. I. Guijman y A. V. Skorojod).

Ejemplo (*diferenciación*). Examinemos *la derivada media cuadrática*

$$\xi'(t) = \text{l.i.m.}_{\Delta t \to \infty} \frac{\xi(t+\Delta t) - \xi(t)}{\Delta t}$$

del proceso aleatorio $\xi(t)$ (se tiene en cuenta el límite medio cuadrático). Para el proceso estacionario $\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda)$ en el caso de existencia de tal derivada obtenemos

$$\xi'(t) = \int_{-\infty}^{\infty} i\lambda e^{i\lambda t}\, d\Phi(\lambda). \tag{7.31}$$

Naturalmente, el límite medio cuadrático indicado

$$\text{l.i.m.}_{\Delta t \to 0} \frac{\xi(t+\Delta t) - \xi(t)}{\Delta t}$$

y la integral estocástica $\int_{-\infty}^{\infty} i\lambda e^{i\lambda t}\, d\Phi(\lambda)$ no siempre existen y, precisamente, entonces y sólo entonces, cuando

$$\int_{-\infty}^{\infty} \lambda^2 f(\lambda)\, d\lambda < \infty,$$

En efecto, para esta condición $\varphi(\lambda) = i\lambda e^{i\lambda t}$ es el límite de la función $\frac{e^{i\lambda(t+\Delta t)} - e^{i\lambda t}}{\Delta t}$, $\Delta t \to 0$, en el sentido de la convergencia media cuadrática y

$$\frac{\xi(t+\Delta t) - \xi(t)}{\Delta t} = \int_{-\infty}^{\infty} \left[\frac{e^{i\lambda(t+\Delta t)} - e^{i\lambda t}}{\Delta t}\right] d\Phi(\lambda) \to \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda).$$

Ejemplo (*fórmula de inversión*). Examinemos la fórmula (7.30) para

$$\varphi(\lambda) = \int_{-\infty}^{\infty} \chi_\Delta(\mu)\, \delta(\lambda - \mu)\, d\mu, \quad -\infty < \lambda < \infty,$$

donde $\Delta = (\lambda_1, \lambda_2]$ es un intervalo finito, $\chi_\Delta(\lambda)$ es su indicador ($\chi_\Delta(\lambda) = 1$ para $\lambda \in \Delta$ y $\chi_\Delta(\lambda) = 0$ para $\lambda \notin \Delta$) y

$$\delta(\lambda) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\lambda^2/2\sigma^2};$$

la función correspondiente $c(t)$ es

$$c(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\lambda t}\varphi(\lambda)\,d\lambda=\frac{1}{2\pi}\,\frac{e^{-i\lambda_2 t}-e^{-i\lambda_1 t}}{-it}\,e^{-\sigma^2 t/2},$$

$-\infty<t<\infty$.

Se ve fácilmente, que para $\sigma\to 0$ las funciones $\varphi(\lambda)$ convergen en la media cuadrática, hacia la función $\chi_\Delta(\lambda)$ (esto se puede fundamentar lo mismo que en la relación límite (7.29), utilizando el hecho de que las funciones $\varphi(\lambda)$ están limitadas y convergen uniformemente hacia $\chi_\Delta(\lambda)$ según λ, para $-a\leqslant\lambda\leqslant\lambda_1-\varepsilon$, $\lambda_1+\varepsilon\leqslant\lambda\leqslant\lambda_2-\varepsilon$, $\lambda_2+\varepsilon\leqslant\lambda\leqslant a$ para cualesquiera valores de a tan grandes como se quieran y de ε tan pequeños como se deseen. Ya que las magnitudes

$$\int_{-\infty}^{\infty}\varphi(\lambda)\,d\Phi(\lambda)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\lambda_2 t}-e^{-i\lambda_1 t}}{-it}\,e^{-\sigma^2 t^2/2}\xi(t)\,dt$$

entran en el espacio $H(T)$ para $T=(-\infty,\infty)$, entonces en este espacio se contienen también las magnitudes $\Delta\Phi=\Phi(\lambda_2)-\Phi(\lambda_1)$:

$$\Delta\Phi=\int_{-\infty}^{\infty}\chi_\Delta(\lambda)\,d\Phi(\lambda)=\underset{\sigma\to 0}{\text{l.i.m.}}\,\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\lambda_2 t}-e^{-i\lambda_1 t}}{-it}\,e^{-\sigma^2 t^2/2}\,\xi(t)\,dt \tag{7.32}$$

Esta es la llamada *fórmula de inversión*, que permite determinar el proceso aleatorio $\Phi(\lambda)$, $-\infty<\lambda<\infty$, con incrementos no correlacionados que figuran en la representación espectral del proceso estacionario $\xi(t)$, según su trayectoria $\xi(t)$, $-\infty<t<\infty$.

Examinemos el sistema lineal con la función de peso $w(t)$ absolutamente integrable

$$w(t)=0 \text{ para } t<0,\quad \int_0^{\infty}|w(t)|^2\,dt<\infty,$$

y con la característica espectral $\varphi(i\lambda)$, $\int_{-\infty}^{\infty}|\varphi(i\lambda)|^2\,d\lambda<\infty$.

Supongamos que en la entrada de este sistema aparece el proceso aleatorio estacionario en amplio sentido

$$\dot{\eta}(t)=\int_{-\infty}^{\infty}e^{i\lambda t}\,d\Psi(\lambda) \tag{7.33}$$

con la densidad espectral $g(\lambda)$. Para tal acción exterior durante el tiempo desde 0 hasta t, el sistema pasa del estado inicial (digamos

$\xi(0) = 0$) al estado

$$\xi(t) = \int_0^t w(t-s)\,\dot{\eta}(s)\,ds. \tag{7.34}$$

De acuerdo con la igualdad general (7.30) el proceso aleatorio $\xi(t)$ se puede representar en la forma

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\varphi_t(\lambda)\,d\Psi(\lambda),$$

donde la función subintegral $\varphi_t(\lambda) = \int_0^t e^{-i\lambda s} w(s)\,ds$ converge hacia la característica espectral $\varphi(i\lambda)$ para $t \to \infty$. Se ve fácilmente que *el proceso aleatorio $\xi(t)$ converge en su media cuadrática hacia el proceso estacionario de la forma*[1])

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\varphi(i\lambda)\,d\Psi(\lambda). \tag{7.35}$$

La fórmula (7.35) nos da de hecho la representación espectral de este proceso estacionario $\xi(t)$, que se establece en el curso del tiempo ($t \to \infty$) a la salida del sistema lineal con la característica espectral $\varphi(i\lambda)$ o sea en la representación espectral

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\,d\Phi(\lambda)$$

es necesario tomar

$$\Phi(\lambda) = \int_{-\infty}^{\lambda} \varphi(i\mu)\,d\Psi(\mu), \quad -\infty < \lambda < \infty.$$

De la relación

$$\mathbf{M}\,|\,\Phi(\lambda_2) - \Phi(\lambda_1)\,|^2 = \int_{\lambda_1}^{\lambda_2} |\,\varphi(i\lambda)\,|^2\,g(\lambda)\,d\lambda$$

se ve directamente que la densidad espectral $f(\lambda)$ del proceso estacionario $\xi(t)$ es

$$f(\lambda) = |\,\varphi(i\lambda)\,|^2\,g(\lambda). \tag{7.36}$$

[1]) Del mismo modo, que en el punto 3 del § 6, se podría examinar el paso límite (para $t_0 \to -\infty$) del proceso $\xi^0(t) = \int_{t_0}^t w(t-s)\,\dot{\eta}(s)\,ds$ al proceso estacionario $\xi(t)$ determinado por la fórmula (7.35).

Recordemos que la densidad espectral $f(\lambda)$ caracteriza la distribución de la energía total del proceso estacionario $\xi(t)=\int_{-\infty}^{\infty} e^{i\lambda t}\,d\Phi(\lambda)$ por las componentes de la forma $\xi_\Delta(t)=\int_{\lambda_1}^{\lambda_2} e^{i\lambda t}\,d\Phi(\lambda)$ en dependencia del intervalo $\Delta=(\lambda_1, \lambda_2)$ de aquellas frecuencias λ que entran en e espectro de frecuencias de la componente correspondiente $\xi_\Delta(t)$ $\left(\text{a saber, } \xi_\Delta(t) \text{ tiene una energía igual a } \int_{\lambda_1}^{\lambda_2} f(\lambda)\,d\lambda\right)$. La relación (7.36) indica que al elegir la característica espectral adecuada $\varphi(i\lambda)$ del sistema lineal, se puede reforzar (o debilitar) las "oscilaciones de entrada" $\eta_\Delta=\int_{\lambda_1}^{\lambda_2} e^{i\lambda t}\,d\Psi(\lambda)$ cuya energía $\int_{\lambda_1}^{\lambda_2} g(\lambda)\,d\lambda$ se transforma en $\xi_\Delta(t)=\int_{\lambda_1}^{\lambda_2} e^{i\lambda t}\varphi(i\lambda)\,d\Psi(\lambda)$; como ya se indicó la energía de las "oscilaciones de salida" $\xi_\Delta(t)=\int_{\lambda_1}^{\lambda_2} e^{i\lambda t}\,d\Phi(\lambda)$ será igual a $\int_{\lambda_1}^{\lambda_2} f(\lambda)\,d\lambda$, donde $f(\lambda)=|\varphi(i\lambda)|^2\,g(\lambda)$.

Anteriormente, en el § 6 examinamos el comportamiento del sistema bajo la acción de perturbaciones aleatorias homogéneas $\dot{\eta}(t)$ con la suposición de que $\eta(t)=\int_0^t \dot{\eta}(s)\,ds$ representa en sí el proceso con incrementos no correlacionados, que satisface a las condiciones (7.14); en estas condiciones el proceso de salida

$$\xi(t)=\int_0^t w(t-s)\,d\eta(s),$$

que se puede representar simbólicamente por la fórmula (7.34) con el cambio de $d\eta(s)$ por $\dot{\eta}(s)\,ds$, será convergente hacia el proceso estacionario para $t\to\infty$. Con respecto a este proceso estacionario

$$\xi(t)=\int_{-\infty}^{\infty} e^{i\lambda t}\varphi(i\lambda)\,d\Psi(\lambda)$$

fue establecido que él admite la representación espectral y tiene una densidad espectral $f(\lambda)=\frac{\sigma^2}{2\pi}|\varphi(i\lambda)|^2$ (véase (7.22)—(7.25)).

Tal proceso estacionario entra formalmente en el esquema general descrito anteriormente cuando $g(\lambda)=\frac{\sigma^2}{2\pi}$. Naturalmente, en el sentido de la definición dada antes, la función $g(\lambda)=\frac{\sigma^2}{2\pi}$ no puede servir como densidad espectral del proceso estacionario corriente $\left(\text{ya que para ella } \int\limits_{-\infty}^{\infty} g(\lambda)\,d\lambda=\infty\right)$. Pero es cómodo, de un modo puramente simbólico, introducir el proceso estacionario correspondiente $\dot{\eta}(t)$ con la densidad espectral $g(\lambda)=\frac{\sigma^2}{2\pi}$, utilizándolo en el sentido de que (véase (7.19))

$$\int\limits_{-\infty}^{\infty} c(t)\,\dot{\eta}(t)\,dt=\left(\int\limits_{-\infty}^{\infty} c(t)\,d\eta(t)\right)=\int\limits_{-\infty}^{\infty}\varphi(\lambda)\,d\Psi(\lambda).$$

Este proceso estacionario generalizado $\eta(t)$ se llama, corrientemente, *ruido blanco*.

Ejemplo (movimiento perturbado del péndulo). En conclusión, examinemos un ejemplo en el cual se puede fácilmente observar algunas particularidades del comportamiento de los sistemas lineales bajo acción de las oscilaciones aleatorias. Examinemos precisamente el movimiento del péndulo (para la existencia del rozamiento) con un gran período de oscilaciones propias, que se describen por la ecuación

$$x''(t)+2hx'(t)+\omega_0^2 x(t)=0$$

y se dan en forma explícita por la función

$$x(t)=Ae^{-ht}\operatorname{sen}(\omega t+\theta),\quad \omega=\sqrt{\omega_0^2-h^2}.$$

La función de transmisión de tal sistema es

$$\varphi(p)=\frac{1}{p^2+2hp+\omega_0^2}$$

(la característica espectral es igual a $\varphi(i\lambda)$).

Supongamos que este sistema se encuentra en un barco. Y supongamos que sea la frecuencia ω de oscilaciones propias del péndulo mucho menor que la frecuencia Ω del balanceo del barco: $\omega \ll \Omega$.

Si consideramos que en el péndulo actúan los golpes aleatorios *como resultado del balanceo, que se originan aproximadamente* después de pequeños intervalos de tiempo $\Delta t\sim\frac{\pi}{\Omega}$, siendo esta perturbación externa $\dot{\eta}(t)$ homogénea según el tiempo y se puede considerar formalmente que $\dot{\eta}(t)$ es «un ruido blanco»[1]), entonces el movimiento

establecido del péndulo representará en sí un proceso estacionario con la densidad espectral de la forma $f(\lambda) = \frac{\sigma^2}{2\pi} \mid \varphi(i\lambda) \mid^2$.

Recordemos que la densidad espectral caracteriza la distribución de energía del proceso aleatorio $\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda)$ según las componentes de «las oscilaciones elementales» en dependencia de la frecuencia λ, a saber: la amplitud media de las oscilaciones aleatorias, dadas por la integral estocástica $\int_{\lambda_1}^{\lambda_2} e^{i\lambda t}\, d\Phi(\lambda)$ es igual a $\int_{\lambda_1}^{\lambda_2} f(\lambda)\, d\lambda$. En nuestro caso la densidad espectral es

$$f(\lambda) = \frac{\sigma^2}{2\pi} \cdot \frac{1}{(\lambda^2 - \omega_0^2) + 4h^2\lambda^2}$$

y tiene un máximo para $\lambda^2 = \omega^2 - h^2$ (fuertemente expresado para «un coeficiente de rozamiento» h pequeño) y decrece suficientemente rápido al alejarse del punto máximo (fig. 23). Esto quiere decir que en el proceso aleatorio $\xi(t)$ predominan, fuertemente, «las oscilaciones elementales» con frecuencias próximas a la frecuencia propia ω del sistema examinado.

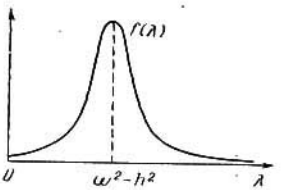

Fig. 23. Vista general de la densidad espectral $f(\lambda) = \frac{\sigma^2}{\pi} \cdot \frac{1}{(\lambda^2 - \omega_0^2)^2 + 4h^2\lambda^2}$.

Para la comparación indiquemos que si se considera la acción externa (del balanceo del barco) como oscilación armónica de frecuencia Ω, entonces el péndulo durante el movimiento establecido realizará oscilaciones forzadas con esta frecuencia Ω, lo que se diferencia cualitativamente del resultado obtenido anteriormente.

8. PROCESOS DE DIFUSION

1. Procesos aleatorios representados por la integral estocástica Itô. Para mayor claridad, imaginémonos un sistema determinado, que con la acción exterior $y(t)$, $t \geqslant t_0$, cambia su estado físico $x(t)$ de tal modo que

$$x'(t) = a(t) + \sigma(t)\, y(t),$$

[1]) Referente a la acción de tal tipo de proceso aleatorio sobre el sistema lineal, además de lo dicho anteriormente sobre «el ruido blanco» véase también el punto 1 del § 6.

o sea en forma integral

$$x(t) = x(t_0) + \int_{t_0}^{t} a(s)\,ds + \int_{t_0}^{t} \sigma(s)\,y(s)\,ds,$$

donde los coeficientes $a(t)$ y $\delta(t)$, que son parámetros del sistema, pueden depender no sólo del tiempo t, sino también de la acción $y(s)$, $t_0 \leqslant s \leqslant t$. Al actuar sobre tal sistema las perturbaciones caóticas aleatorias $\dot{\eta}(t)$, $t \geqslant t_0$, como modelo adecuado que describe su comportamiento, puede servir el proceso aleatorio

$$\xi(t) = \xi(t_0) + \int_{t_0}^{t} a(s)\,ds + \int_{t_0}^{t} \sigma(s)\,d\eta(s), \tag{8.1}$$

donde $\eta(t) = \int_{t_0}^{t} \dot{\eta}(t)\,dt$ (compárese el punto 1 del § 6).

Más adelante examinaremos los procesos aleatorios $\xi(t)$ del tipo indicado, suponiendo que $\eta(t)$, $t \geqslant t_0$, es el proceso «standard» del movimiento browniano[1]) con una media nula y el coeficiente unitario de difusión ($\mathbf{M}[\eta(t) - \eta(t_0)]^2 = t - t_0$), y los coeficientes $a(t)$, $\sigma(t)$ dependen, hablando en general, de $\eta(s)$, $t_0 \leqslant s \leqslant t$ (más exacto, $a(t)$ y $\delta(t)$ para cada t son magnitudes aleatorias, cuyos valores en el momento t se determinan unívocamente por las trayectorias del proceso aleatorio $\eta(s)$, $t_0 \leqslant s \leqslant t$).

Pasemos a la definición exacta de las integrales estocásticas existentes en la fórmula (8.1).

Sea $\varphi(t)$ una función real constante a trozos en el segmento $c_1 \leqslant t \leqslant c_2$, que para una división de este segmento por los puntos $c_1 = t_0, t_1, \ldots, t_n = c_2$ toma los valores aleatorios $\varphi(t_k)$ en cada uno de los intervalos $\Delta_k = (t_k, t_{k+1})$, $k = 0, \ldots, n-1$. Subrayamos una vez más que los valores de la función $\varphi(t)$ son magnitudes aleatorias. Supongamos que

$$\int_{c_1}^{c_2} \varphi(t)\,dt = \sum_{k=0}^{n-1} \varphi(t_k)\,\Delta t_k,$$

donde $\Delta t_k = t_{k+1} - t_k$, $k = 0, \ldots, n-1$. Está claro, que la integral definida de este modo posee las propiedades corrientes de aditividad y homogeneidad:

$$\int_{c_1}^{c_2} [\lambda_1 \varphi_1(t) + \lambda_2 \varphi_2(t)]\,dt = \lambda_1 \int_{c_1}^{c_2} \varphi_1(t)\,dt + \lambda_2 \int_{c_1}^{c_2} \varphi_2(t)\,dt \tag{8.2}$$

[1]) En otras palabras, $\eta(t)$ es el «ruido blanco» de Gauss.

y ademas de esto,

$$\mathbf{M}\int_{c_1}^{c_2}\varphi(t)\,dt=\int_{c_1}^{c_2}[\mathbf{M}\varphi(t)]\,dt \tag{8.3}$$

y

$$\left\|\int_{c_1}^{c_2}\varphi(t)\,dt\right\|\leqslant\int_{c_1}^{c_2}\|\varphi(t)\|\,dt \tag{8.4}$$

(donde $\|\eta\|=\sqrt{\mathbf{M}\eta^2}$ significa el valor medio cuadrático de la magnitud aleatoria η).

Sea ahora $\varphi(t)$ una función aleatoria en el segmento $c_1\leqslant t\leqslant c_2$, que puede ser representada como límite medio cuadrático de las funciones constantes a trozos $\varphi_n(t)$, $n=1, 2, \ldots$, o más exáctamente,

$$\|\varphi(t)-\varphi_n(t)\|\to 0 \quad \text{para} \quad n\to\infty \tag{8.5}$$

uniformemente según t, $c_1\leqslant t\leqslant c_2$. Evidentemente,

$$\|\varphi_n(t)-\varphi_m(t)\|\leqslant\|\varphi_n(t)-\varphi(t)\|+\|\varphi_m(t)-\varphi(t)\|\to 0$$

para n, $m\to\infty$ también uniformemente según t y, por consiguiente,

$$\left\|\int_{c_1}^{c_2}\varphi_n(t)\,dt-\int_{c_1}^{c_2}\varphi_m(t)\,dt\right\|\leqslant\int_{c_1}^{c_2}\|\varphi_n(t)-\varphi_m(t)\|\,dt\to 0.$$

Esto demuestra que existe el límite medio cuadrático $\text{l.i.m.}_{n\to\infty}\int_{c_1}^{c_2}\varphi_n(t)\,dt$; evidentemente, este límite es el mismo también para cualquiera sucesión $\varphi_n(t)$, $n=1, 2, \ldots$, del tipo indicado anteriormente. Supongamos que

$$\int_{c_1}^{c_2}\varphi(t)\,dt=\text{l.i.m.}_{n\to\infty}\int_{c_1}^{c_2}\varphi_n(t)\,dt. \tag{8.6}$$

Se ve fácilmente, que en tal paso límite se conservan las relaciones (8.2)—(8.4).

Señalemos que para la función aleatoria $\varphi(t)$, *uniformemente continua en la media cuadrática*:

$$\|\varphi(t+\Delta t)-\varphi(t)\|\to 0 \tag{8.7}$$

para $\Delta t\to 0$ uniformemente según t, $c_1\leqslant t\leqslant c_2$; la *integral estocástica* determinada por la igualdad (8.6) es el límite medio cuadrático de las sumas integrales corrientes:

$$\int_{c_1}^{c_2}\varphi(t)\,dt=\text{l.i.m.}\sum_{k=0}^{n-1}\varphi(t_k)\,\Delta t_k$$

para una división cada vez más pequeña del segmento $[c_1, c_2]$ por los puntos $c_1 = t_0, t_1, \ldots, t_n = c_2$ en intervalos de longitud $\Delta t_k = t_{k+1} - t_k$, cuando máx $\Delta t_k \to 0$.

Pasemos a la definición de integral estocástica del tipo $\int_{c_1}^{c_2} \varphi(t) \times d\eta(t)$, donde $\varphi(t)$ es una función aleatoria, y $\eta(t)$ un proceso aleatorio con incrementos independientes; con esto nos limitamos solamente al caso cuando el proceso $\eta(t)$ es el movimiento browniano (con una media nula y con el coeficiente unitario de difusión).

Para las funciones corrientes $\varphi(t)$ (no aleatorias), la integral estocástica de tal tipo fue determinada antes en el § 6. Para las funciones aleatorias $\varphi(t)$ es fundamental la suposición de que

para cualquier u los incrementos $\eta(t) - \eta(u)$, $t \geqslant u$, no dependen de las magnitudes aleatorias $\varphi(s)$, $s \leqslant u$.

Sea $\varphi(t)$ una función constante a trozos real, que toma los valores aleatorios $\varphi(t_k)$ en los intervalos correspondientes $\Delta_k = (t_k, t_{k+1})$, $k = 0, \ldots, n-1$. Pongamos

$$\int_{c_1}^{c_2} \varphi(t)\, d\eta(t) = \sum_{k=0}^{n-1} \varphi(t_k)\, \Delta\eta(t_k),$$

donde $\Delta\eta(t_k) = \eta(t_{k+1}) - \eta(t_k)$; $k = 0, \ldots, n-1$. Está claro, que la integral así determinada posee las propiedades corrientes de aditividad y homogeneidad (compárese (8.2)).

Supondremos que $\mathbf{M}[\varphi(t)]^2 < \infty$ para todos los t, $c_1 \leqslant t \leqslant c_2$. De la condición de independencia de los incrementos $\eta(t) - \eta(u)$, $t \geqslant u$, de los valores $\varphi(s)$, $s \leqslant u$, se deduce que

$$\mathbf{M}[\varphi(t_k) \cdot \Delta\eta(t_k)] = \mathbf{M}\varphi(t_k) \cdot \mathbf{M}\Delta\eta(t_k) = 0,$$

$$\mathbf{M}[\varphi(t_k) \cdot \Delta\eta(t_k)]^2 = \mathbf{M}\varphi^2(t_k) \cdot \mathbf{M}[\Delta\eta(t_k)]^2 = \mathbf{M}\varphi^2(t_k) \cdot \Delta(t_k),$$

$$k = 0, \ldots, n-1;$$

además de esto, para todos los $k \neq j$ (considerando para concretizar que $k < j$) tenemos

$$\mathbf{M}[\varphi(t_k)\, \Delta\eta(t_k) \cdot \varphi(t_j)\, \Delta\eta(t_j)] = \mathbf{M}[\varphi(t_k)\, \Delta\eta(t_k)\, \varphi(t_j)] \times \mathbf{M}\Delta\eta(t_j) = 0.$$

Evidentemente,

$$\mathbf{M} \int_{c_1}^{c_2} \varphi(t)\, d\eta(t) = 0, \tag{8.8}$$

$$\left\| \int_{c_1}^{c_2} \varphi(t)\, d\eta(t) \right\|^2 = \int_{c_1}^{c_2} \| \varphi(t) \|^2\, dt$$

y además de esto,

$$\mathbf{M}\left[\int_{c_1}^{c_2}\varphi(t)\,d\eta(t)\cdot\int_{c_1}^{c_2}\psi(t)\,d\eta(t)\right]=\int_{c_1}^{c_2}[\mathbf{M}\,\varphi(t)\,\psi(t)]\,dt. \qquad (8.9)$$

Sea ahora $\varphi(t)$ una función aleatoria en el segmento $c_1 \leqslant t \leqslant c_2$, que puede ser representada como el límite medio cuadrático de las funciones constantes a trozos $\varphi_n(t)$, $n = 1, 2, \ldots$; más exactamente, supongamos cumplida la condición (8.5). Entonces

$$\left\|\int_{c_1}^{c_2}\varphi_n(t)\,d\eta(t)-\int_{c_1}^{c_2}\varphi_m(t)\,d\eta(t)\right\|^2=\int_{c_1}^{c_2}\|\varphi_n(t)-\varphi_m(t)\|^2dt\to 0$$

para $n, m \to \infty$ y, por consiguiente, existe el límite medio cuadrático $\underset{n\to\infty}{\text{l.i.m.}}\int_{c_1}^{c_2}\varphi_n(t)\,d\eta(t)$. Se ve fácilmente, que este límite no depende de la elección de la sucesión $\varphi_n(t)$, $n = 1, 2, \ldots$, que satisfaga la condición (8.5). Pongamos

$$\int_{c_1}^{c_2}\varphi(t)\,d\eta(t)=\underset{n\to\infty}{\text{l.i.m.}}\int_{c_1}^{c_2}\varphi_n(t)\,d\eta(t). \qquad (8.10)$$

Evidentemente, para tal paso límite, las relaciones (8.8), (8.9) se extienden a las funciones límites $\varphi(t)$.

Señalemos, que para las funciones uniformemente continuas $\varphi(t)$ (en la media cuadrática) la integral estocástica definida por la igualdad (8.10) es el límite medio cuadrático de las sumas integrales correspondientes:

$$\int_{c_1}^{c_2}\varphi(t)\,d\eta(t)=\text{l.i.m.}\sum_{k=0}^{n-1}\varphi(t_k)\,\Delta\eta(t_k).$$

Después de introducir las integrales estocásticas del tipo (8.6), (8.10) también se puede hablar con conocimiento de causa sobre el proceso aleatorio $\xi(t)$ representado por la fórmula (8.1).

Suponemos que las funciones aleatorias $a(t)$ y $\sigma(t)$ cuyos valores para cada t se determinan unívocamente por el proceso aleatorio $\eta(s)$, $t_0 \leqslant s \leqslant t$, son uniformemente continuas en la media cuadrática en cada segmento limitado. En esta suposición existe la integral $\int_{t_0}^{t}a(s)\,ds$, y ya que para cualquier u, los incrementos $\eta(t)-\eta(u)$, $t \geqslant u$, no dependen de $\eta(s)$, $t_0 \leqslant s \leqslant u$, ellos tampoco dependen de las magnitudes $\sigma(s)$, $s \leqslant u$

de modo que existe la integral estocástica $\int_{t_0}^{t} \sigma(t)\, d\eta\ (t)$ determinada anteriormente.

Examinemos el caso en que el proceso aleatorio $\xi(t)$ es representable por la integral estocástica (8.1). Aquí consideramos que existen los procesos aleatorios $a(t)$, $\sigma(t)$ y $\eta(t)$ (y también el proceso aleatorio $\xi(t)$) y la cuestión se concierne sólo al enlace de estos procesos, expresado por la fórmula (8.1).

Claro está que los valores $\xi(t)$ dados por esta fórmula se determinan unívocamente por el valor inicial $\xi(t_0)$ y por el proceso aleatorio $\eta(s)$, $t_0 \leqslant s \leqslant t$; examinaremos el proceso aleatorio $\xi(t)$, $t \geqslant t_0$ de este tipo. Se indicará que la relación integral (8.1) es un corolario de la relación local

$$\Delta\xi(t) \sim a(t)\,\Delta t + \sigma(t)\,\Delta\eta(t) \tag{8.11}$$

(donde $\Delta\xi(t) = \xi(t+\Delta t) - \xi(t)$, $\Delta\eta(t) = \eta(t+\Delta t) - \eta(t)$ y $\Delta t \to 0$), que comprenderemos en el sentido[1]) de que

$$\mathbf{M}\{\Delta\xi(t) - [a(t)\,\Delta t + \sigma(t)\,\Delta\eta(t)] \mid \eta(s),\ s \leqslant t\} = o(\Delta t), \tag{8.12}$$
$$\mathbf{M}\{\Delta\xi(t) - [a(t)\,\Delta t + \sigma(t)\,\Delta\eta(t)]\}^2 = o(\Delta t),$$

donde las magnitudes $o(\Delta t)$ son infinitamente pequeñas de orden superior en comparación con Δt, o más exacto,

$$\frac{\|o(\Delta t)\|}{\Delta t} \to 0 \quad \text{para} \quad \Delta t \to 0$$

uniformemente por t en cada intervalo finito $t_0 \leqslant t \leqslant t_1$.

Detengámonos detalladamente en estas relaciones. Las magnitudes $a(t)$ y $\sigma(t)$ en la primera de éstas se determinan unívocamente por $\eta(s)$, $s \leqslant t$, y para la trayectoria fijada $\eta(s)$, $s \leqslant t$, son sencillamente constantes, mientras que el incremento $\Delta\xi(t)$ es aleatorio y tiene una distribución de probabilidades que depende de la trayectoria correspondiente $\eta(s)$, $s \leqslant t$, y el incremento $\Delta\eta(t)$ no depende de $\eta(s)$, $s \leqslant t$. Por eso, para la esperanza matemática condicional (para los valores fijados de $\eta(s)$, $s \leqslant t$) tenemos

$$\mathbf{M}\{\sigma(t)\,\Delta\eta(t) \mid \eta(s),\ s \leqslant t\} = \sigma(t)\,\mathbf{M}\Delta\eta(t) = 0,$$
$$\mathbf{M}\{\Delta\xi(t) - [a(t)\,\Delta t + \sigma(t)\,\Delta\eta(t)] \mid \eta(s),\ s \leqslant t\} =$$
$$= \mathbf{M}\{\Delta\xi(t) \mid \eta(s),\ s \leqslant t\} - a(t)\,\Delta t.$$

[1]) Véase el § 2 del siguiente cap. IV. Más adelante, en este párrafo utilizaremos, esencialmente, el aparato de las esperanzas matemáticas condicionales, incluyendo la fórmula de la esperanza matemática repetida del tipo

$$\mathbf{M}\{\mathbf{M}[\xi \mid \eta] \mid \zeta\} = \mathbf{M}\{\xi \mid \zeta\},$$

donde, hablando a grueso modo, el complejo de condiciones «ζ» se determina unívocamente para cada una de las condiciones fijas «η»; se puede leer más detalladamente sobre este particular, por ejemplo, en el libro citado antes de Guijman y Skorojod.

Luego, ya que se supone que la función aleatoria $a(t)$ es continua en la media cuadrática, la magnitud $\| a(t) \|$ está limitada en cada segmento $t_0 \leqslant t \leqslant t_1$ y, se ve fácilmente, que

$$\| \Delta\xi(t) - [a(t)\,\Delta t + \sigma(t)\,\Delta\eta(t)] \|^2 = \\ = \| \Delta\xi(t) - \sigma(t)\,\Delta\eta(t) \|^2 + o(\Delta t).$$

De tal modo, las relaciones (8.12) también se pueden comprender en el sentido de que

$$\mathbf{M}\{\Delta\xi(t) \mid \eta(s),\ s \leqslant t\} = a(t)\,\Delta t + o(\Delta t), \tag{8.13}$$

$$\| \Delta\xi(t) - \sigma(t)\,\Delta\eta(t) \|^2 = o(\Delta t). \tag{8.14}$$

Lema 1. *Con la condición* (8.13) *para cualesquiera* $t_1 \leqslant t_2$ *tiene lugar la siguiente igualdad:*

$$\mathbf{M}\{\xi(t_2) - \xi_1(t_1) \mid \eta(s),\ s \leqslant t_1\} = \mathbf{M}\left\{\int_{t_1}^{t_2} a(t)\,dt \mid \eta(s),\ s \leqslant t_1\right\}. \tag{8.15}$$

Demostración. Dividamos el segmento $[t_1, t_2]$ con los puntos $t_1 = s_0, s_1, \ldots, s_n = t_2$ en intervalos de longitud $\Delta s_k = s_{k+1} - s_k$, $k = 0, \ldots, n-1$. Supongamos que

$$o(\Delta s_k) = \mathbf{M}\{\Delta\xi(s_k) - a(s_k)\,\Delta s_k \mid \eta(s),\ s \leqslant s_k\},$$

donde $\Delta\xi(s_k) = \xi(s_{k+1}) - \xi(s_k)$; $k = 0, \ldots, n-1$. Según la suposición, $\frac{\| o(\Delta s_k) \|}{\Delta s_k} \to 0$ uniformemente por k para máx $\Delta s_k \to 0$. Tomando por segunda vez la esperanza matemática condicional (pero ahora con la condición $\eta(s)$, $s \leqslant t_1$, donde $t_1 \leqslant s_k$) tenemos

$$\mathbf{M}\{o(\Delta s_k) \mid \eta(s),\ s \leqslant t_1\} = \mathbf{M}\{\Delta\xi(s_k) - \\ - a(s_k)\,\Delta s_k \mid \eta(s),\ s \leqslant t_1\},$$

de donde obtenemos que

$$\| \mathbf{M}\{\Delta\xi(s_k) - a(s_k)\,\Delta s_k \mid \eta(s),\ s \leqslant t_1\} \| = \\ = \sqrt{\mathbf{M}[\mathbf{M}\{o(\Delta s_k) \mid \eta(s),\ s \leqslant t_1\}]^2} \leqslant \\ \leqslant \sqrt{\mathbf{M}[\mathbf{M}\{o(\Delta s_k)^2 \mid \eta(s),\ s \leqslant t_1\}]} = \\ = \| o(\Delta s_k) \|,\quad k = 0, \ldots, n-1.$$

Por esto, como se ve fácilmente,

$$\left\| \mathbf{M}\{\xi(t_2) - \xi(t_1) \mid \eta(s),\ s \leqslant t_1\} - \mathbf{M}\left\{\int_{t_1}^{t_2} a(t)\,dt \mid \eta(s),\ s \leqslant t_1\right\} \right\| \leqslant \\ \leqslant \sum_{k=0}^{n-1} \mathbf{M}\{\Delta\xi(s_k) - a(s_k)\,\Delta s_k \mid \eta(s),\ s \leqslant t_1\} \| +$$

$$+\left\|\mathbf{M}\left\{\int_{t_1}^{t_2} a(t)\,dt-\sum_{k=0}^{n-1} a(s_k)\,\Delta s_k \mid \eta(s),\ s\leqslant t_1\right\}\right\|\leqslant$$

$$\leqslant\sum_{k=0}^{n-1}\|o(\Delta s_k)\|+\left\|\int_{t_1}^{t_2} a(t)\,dt-\sum_{k=0}^{n=1} a(s_k)\,\Delta s_k\right\|\to 0$$

para máx $\Delta s_k \to 0$ y, por consiguiente,

$$\|\mathbf{M}\{\xi(t_2)-\xi(t_1)\mid\eta(s),\ s\leqslant t_1\}-\mathbf{M}\left\{\int_{t_1}^{t_2} a(t)\,dt\mid\eta(s),\ s\leqslant t_1\right\}\|=0,$$

es decir, tiene lugar la igualdad (8.15).

Pasemos del proceso inicial $\xi(t)$ al proceso aleatorio

$$\tilde{\xi}(t)=\xi(t)-\xi(t_0)-\int_{t_0}^{t} a(s)\,ds. \tag{8.16}$$

De la igualdad (8.15) se deduce que

$$\mathbf{M}\{\Delta\tilde{\xi}(t)\mid\eta(s),\quad s\leqslant t\}=0. \tag{8.17}$$

Además de esto, junto con la condición (8.14) para los incremento

$$\Delta\tilde{\xi}(t)=\Delta\xi(t)-\int_{t}^{t+\Delta t} a(s)\,ds$$

será evidentemente

$$\|\Delta\tilde{\xi}(t)-\sigma(t)\,\Delta\eta(t)\|^2=o(\Delta t), \tag{8.18}$$

donde $o(\Delta t)$ es un infinitamente pequeño del mismo tipo que en (8.14), es decir, $\frac{o(\Delta t)}{\Delta t}\to 0$ uniformemente en cada intervalo finito de $t_0\leqslant t\leqslant t_1$.

Lema 2. *Con la condición* (8.14) *tiene lugar la siguiente igualdad*:

$$\tilde{\xi}(t)=\int_{t_0}^{t}\sigma(s)\,d\eta(s). \tag{8.19}$$

Demostración. Elijamos una división del segmento $[t_0,\ t]$ y valoremos la diferencia correspondiente

$$\tilde{\xi}(t)-\sum_{k=0}^{n-1}\sigma(t_k)\,\Delta\eta(t_k)=\sum_{k=0}^{n-1}[\Delta\tilde{\xi}(t_k)-\sigma(t_k)\,\Delta\eta(t_k)].$$

De la condición (8.17) se deduce que para $t_k \neq t_j$ (digamos, $t_k < t_j$)

$$\mathbf{M}\Delta\widetilde{\xi}(t_k)\,\Delta\widetilde{\xi}(t_j) = \mathbf{M}\,[\mathbf{M}\,\{\Delta\widetilde{\xi}(t_k)\,\Delta\widetilde{\xi}(t_j) \mid \eta(s),$$
$$s \leqslant t_j\}] = \mathbf{M}\,[\Delta\widetilde{\xi}(t_k)\,\mathbf{M}\,\{\Delta\widetilde{\xi}(t_j) \mid \eta(s),\ s \leqslant t_j\}] = 0,$$

ya que para los valores fijados $\eta(s)$, $s \leqslant t_j$, las magnitudes $\Delta\widetilde{\xi}(t_k)$ son constantes, análogamente:

$$\mathbf{M}\,[\Delta\widetilde{\xi}(t_k)\,\sigma(t_j)\,\Delta\eta(t_j)] = 0,$$
$$\mathbf{M}\,[\sigma(t_k)\,\Delta\eta(t_k)\,\sigma(t_j)\,\Delta\eta(t_j)] = 0.$$

De aquí, tomando en consideración la condición (8.18), obtenemos que

$$\left\|\widetilde{\xi}(t) - \sum_{k=0}^{n-1}\sigma(t_k)\,\Delta\eta(t_k)\right\|^2 = \sum_{k=0}^{n-1}\|\Delta\widetilde{\xi}(t_k) - \sigma(t_k)\,\Delta\eta(t_k)\|^2 =$$
$$= \sum_{k=1}^{n-1} o(\Delta t_k) \to 0$$

para máx $\Delta t_k \to 0$, es decir, "las sumas integrales" $\sum_{k=0}^{n-1}\sigma(t_k)\,\Delta\eta(t_k)$ tienen su límite $\widetilde{\xi}(t)$ y, por consiguiente, $\widetilde{\xi}(t) = \int_0^t \sigma(s)\,d\eta(s)$ que es, lo que se quería demostrar.

La expresión

$$d\xi(t) = a(t)\,dt + \sigma(t)\,d\eta(t) \tag{8.20}$$

se llama *diferencial estocástica* del proceso aleatorio $\xi(t)$, si

$$\xi(t) = \xi(t_0) + \int_{t_0}^{t} a(s)\,ds + \int_{t_0}^{t}\sigma(s)\,d\eta(s).$$

Uniendo los lemas 1, 2 obtenemos el resultado siguiente.

Teorema 1. *Con la condición* (8.12) (*equivalente a* (8.13), (8.14)) *el proceso aleatorio* $\xi(t)$ *tiene la diferencial estocástica, expresada por la fórmula* (8.20).

Hallemos la diferencial estocástica de la función aleatoria

$$\xi(t) = \varphi(t,\ \eta(t)), \tag{8.21}$$

donde $\eta(t)$ es el proceso del movimiento browniano, y $\varphi(t, x)$ es una función real (no aleatoria) de las variables t y x, que tiene derivadas continuas $\frac{\partial\varphi}{\partial t}$, $\frac{\partial\varphi}{\partial x}$ y $\frac{\partial^2\varphi}{\partial x^2}$. Para que sea más claro el objeto de nuestros cálculos, exponemos inmediatamente la solución:

$$d\xi(t) = \left[\frac{\partial\varphi(t,\ \eta(t))}{\partial t} + \frac{1}{2}\,\frac{\partial^2\varphi(t,\ \eta(t))}{\partial x^2}\right]dt + \frac{\partial\varphi(t,\ \eta(t))}{\partial x}\,d\eta(t). \tag{8.22}$$

Subrayemos aquí particularidad en comparación con la diferencial corriente para las funciones (no aleatorias) diferenciables $\xi(t)$ y $\eta(t)$, que tiene la forma $d\xi(t) = \frac{\partial\varphi(t, \eta(t))}{\partial t}dt + \frac{\partial\varphi(t, \eta(t))}{\partial x}d\eta(t)$; precisamente, en la fórmula (8.22) se contiene el miembro complementario $\frac{1}{2}\frac{\partial^2\varphi(t, \eta(t))}{\partial x^2}dt$, que aparece, como podemos cerciorarnos más adelante, gracias a que, para el movimiento browniano $\eta(t)$ examinado con el coeficiente unitario de difusión $\mathbf{M}[\Delta\eta(t)]^2 = \Delta t$. En general [1]),

$$\mathbf{M}[\Delta\eta(t)]^{2k} = \frac{(2k)!}{k!\,2^k}(\Delta t)^k, \quad k = 1, 2, \ldots, \tag{8.23}$$

y esto aporta su específica al valorar el miembro residual

$$o(\Delta t) = \left[\frac{\partial\varphi(t+\theta_1\Delta t, \eta(t)+\Delta\eta(t)}{\partial t} - \frac{\partial\varphi(t, \eta(t))}{\partial t}\right] + \\ + \frac{1}{2}\left[\frac{\partial^2\varphi(t, \eta(t)+\theta_2\Delta\eta(t))}{\partial x^2} - \frac{\partial^2\varphi(t, \eta(t))}{\partial x^2}\right] \tag{8.24}$$

(donde $|\theta_1|, |\theta_2| \leqslant 1$) en la relación

$$\Delta\xi(t) = \frac{\partial\varphi(t, \eta(t))}{\partial t}\Delta t + \frac{\partial\varphi(t, \eta(t))}{\partial x}\Delta\eta(t) + \\ + \frac{1}{2}\frac{\partial^2\varphi(t, \eta(t))}{\partial x^2}[\Delta\eta(t)]^2 + o(\Delta t), \tag{8.25}$$

que se obtiene en la fórmula de Taylor para las funciones $\varphi(t, x)$.

Supongamos complementariamente, que las derivadas $\frac{\partial\varphi}{\partial t}$, $\frac{\partial\varphi}{\partial x}$ y $\frac{\partial^2\varphi}{\partial x^2}$ satisfacen la tal llamada condición de Lipshitz (para la función $y(t, x)$ ésta significa que

$$|y(t+\Delta t, x+\Delta x) - y(t, x)| \leqslant C_1|\Delta t| + C_2|\Delta x|,$$

donde C_1, C_2 son determinadas constantes). Entonces las funciones aleatorias $\frac{\partial\varphi(t, \eta(t))}{\partial t}$, $\frac{\partial\varphi(t, \eta(t))}{\partial x}$ y $\frac{\partial^2\varphi(t, \eta(t))}{\partial x^2}$ son uniformemente continuas en la media cuadrática y, por consiguiente, las funciones $\left\|\frac{\partial\varphi}{\partial t}\right\|$, $\left\|\frac{\partial\varphi}{\partial x}\right\|$ y $\left\|\frac{\partial^2\varphi}{\partial x^2}\right\|$ están uniformemente limitadas en cada segmento finito $t_0 \leqslant t \leqslant t_1$. Teniendo en cuenta esto, y también

[1]) Para la magnitud de Gauss η con la media nula y la dispersión σ^2, que tiene la función característica $\varphi(u) = e^{-\sigma^2u^2/2}$, $-\infty < u < \infty$, la fórmula $\varphi^{(k)}(0) = i^{-k}\mathbf{M}\eta^k$, $k = 0, 1, \ldots$, nos da

$$\mathbf{M}\eta^{2k} = \frac{(2k)!}{k!\,2^k}\sigma^{2k}, \quad k = 0, 1, \ldots$$

circunstancia de que los incrementos $\Delta\eta(t)$ no dependen de $\eta(s)$, $s \leqslant t$, se puede deducir fácilmente, de las relaciones (8.23)—(8.25), que las funciones

$$a(t) = \frac{\partial \varphi(t, \eta(t))}{\partial t} + \frac{1}{2} \frac{\partial^2 \varphi(t, \eta(t))}{\partial x^2} \quad \text{y} \quad \sigma(t) = \frac{\partial \varphi(t, \eta(t))}{\partial t}$$

satisfacen a las condiciones (8.12) siendo justa, para las magnitudes $o(\Delta t)$ que figuran en estas condiciones en cada segmento limitado $t_0 \leqslant t \leqslant t_1$, la valuación $| o(\Delta t) | \leqslant C(\Delta t)^{3/2}$, donde C es una constante determinada. Según el teorema 1, el proceso aleatorio examinado $\xi(t) = \varphi(t, \eta(t))$ tiene la diferencial estocástica $d\xi(t) = a(t)\,dt + \sigma(t)\,d\eta(t)$.

Ejemplo. Calculemos la integral $\int_{t_0}^{t} \eta(s)\,d\eta(s)$, donde $\eta(t)$, $t \geqslant t_0$, es el proceso del movimiento browniano. Intentemos hallar el proceso aleatorio de la forma $\xi(t) = \varphi(t, \eta(t))$ que tiene la diferencial estocástica $d\xi(t) = \eta(t)\,d\eta(t)$; para tal proceso $\int_{t_0}^{t} \eta(s)\,d\eta(s) = \xi(t) - \xi(t_0)$.

De acuerdo con la fórmula general (8.22), la función incógnita $\varphi(t, x)$ deberá satisfacer las siguientes condiciones:

$$\frac{\partial \varphi}{\partial t} + \frac{1}{2} \frac{\partial^2 \varphi}{\partial x^2} = 0, \quad \frac{\partial \varphi}{\partial x} = x.$$

Hallamos fácilmente que

$$\varphi(t, x) = \frac{1}{2} x^2 + c(t) \quad \text{y} \quad c'(t) + \frac{1}{2} = 0,$$

de donde $\varphi(t, x) = \frac{1}{2}(x^2 - t)$. Como resultado obtenemos

$$\int_{t_0}^{t} \eta(s)\,d\eta(s) = \frac{1}{2}[\eta^2(t) - \eta^2(t_0) - (t - t_0)].$$

Señalemos de modo aleccionador que para la función diferenciable corriente $\eta(t)$, la integral correspondiente es

$$\int_{t_0}^{t} \eta(t)\,d\eta(t) = \frac{1}{2}[\eta^2(t) - \eta^2(t_0)].$$

Examinemos el proceso aleatorio de Márkov $\xi(t)$, que tiene la diferencial estocástica $a(t)\,dt + \sigma(t)\,d\eta(t)$.

Consideraremos que los valores de las funciones aleatorias $a(t)$ y $\sigma(t)$ dependen explícitamente, sólo de t y del estado variable $\xi(t)$,

de modo que $a(t) = a(t, \xi(t))$ y $\sigma(t) = \sigma(t, \xi(t))$, donde $a(t, x)$ y $\sigma(t, x)$ son unas funciones (no aleatorias) del par de variables t, x. En términos generales, esto significa que durante el tiempo Δt el proceso pasa desde cada estado $\xi(t) = x$ al estado $x + \Delta\xi(t)$, donde

$$\Delta\xi(t) \sim a(t, \xi(t))\,\Delta t + \sigma(t, \xi(t))\,\Delta\eta(t).$$

Antes interpretábamos esta relación en el sentido de que se cumpliesen las condiciones (8.12). Reforcemos algo a estas condiciones, sustituyéndolas por las siguientes:

$$\mathbf{M}\{\Delta\xi(t) - [a(t, \xi(t))\,\Delta t + \sigma(t, \xi(t))\,\Delta\eta(t)] \mid \xi(t)\} = o(\Delta t), \quad (8.26)$$
$$\mathbf{M}\{[\Delta\xi(t) - a(t, \xi(t))\,\Delta t - \sigma(t, \xi(t))\,\Delta\eta(t)]^2 \mid \xi(t)\} = o(\Delta t),$$

donde $o(\Delta t)$, igual que anteriormente, significa también para cualquier valor fijado $\xi(t) = x$, una magnitud infinitamente pequeña de orden superior en comparación con $\Delta t \to 0$, uniformemente según t en cada intervalo limitado $t_0 \leqslant t \leqslant t_1$

$$|a(t, x)| \leqslant C, \quad C_1 \leqslant |\sigma(t, x)| \leqslant C_2$$

para las constantes determinadas C, C_1, C_2.

Mostremos que *para la condición* (8.26) *el proceso aleatorio de Márkov* $\xi(t)$ *entra en la clase de los llamados procesos de difusión*, cuyas distribuciones de probabilidades están íntimamente ligadas con las ecuaciones diferenciales, que describen los procesos físicos de difusión[1]) (las ecuaciones mencionadas serán examinadas en el punto siguiente).

Demos una difinición formal del proceso de difusión. Para esto nos será cómodo introducir las siguientes designaciones:

$$\eta_\varepsilon = \begin{cases} \eta \text{ para } |\eta| \leqslant \varepsilon \\ 0 \text{ para } |\eta| > \varepsilon. \end{cases}$$

Al proceso aleatorio de Márkov $\xi(t)$ se le llama proceso *de difusión*, si para cualquier $\varepsilon > 0$, los incrementos $\Delta\xi(t)$ satisfacen las condiciones

$$\left.\begin{aligned} &\mathbf{P}\{|\Delta\xi(t)| > \varepsilon \mid \xi(t) = x\} = o(\Delta t), \\ &\mathbf{M}\{(\Delta\xi(t))_\varepsilon \mid \xi(t) = x\} = a(t, x)\,\Delta t + o(\Delta t), \\ &\mathbf{M}\{(\Delta\xi(t))_\varepsilon^2 \mid \xi(t) = x\} = \sigma^2(t, x)\,\Delta t + o(\Delta t), \end{aligned}\right\} \quad (8.27)$$

donde la magnitud $o(\Delta t)$ tiene el mismo sentido que en las condiciones (8.26).

Señalemos, que si se tienen en cuenta no las magnitudes $(\Delta\xi(t))_\varepsilon$ «cortadas hasta ε», sino los propios incrementos $\Delta\xi(t)$, entonces de las relaciones (8.26) se deduce inmediatamente que

$$\mathbf{M}\{\Delta\xi(t) \mid \xi(t) = x\} = a(t, x)\,\Delta t + o(\Delta t)$$

[1]) Con respecto a esta cuestión véase por ejemplo, el libro de (A. Ya. Jinchin, «Leyes asimptóticas de la teoría de probabilidades»).

y

$$\mathbf{M}\{\Delta\xi_\varepsilon^2(t) \mid \xi(t) = x\} = \sigma^2(t, x)\,\Delta t + o(\Delta t).$$

Para pasar a las magnitudes «cortadas» $(\Delta\xi(t))_\varepsilon$ que figuran en las condiciones (8.27), tenemos que demostrar algunas proposiciones auxiliares.

Más adelante se tratará de las distribuciones condicionales de probabilidades y de las esperanzas matemáticas para los valores fijados $\xi(t) = x$.

Examinemos la magnitud $\eta = a(t, x)\,\Delta t + \sigma(t, x)\,\Delta\eta(t)$, donde $\Delta\eta(t)$ es el incremento del movimiento browniano correspondiente durante el tiempo Δt. La densidad de probabilidad de esta magnitud es $p(y) = \frac{1}{2\sqrt{2\pi\sigma}} e^{-(y-a)^2/2\sigma^2}$, donde para abreviar suponemos $a = a(t, x)\,\Delta t$ y $\sigma^2 = \sigma^2(t, x)\,\Delta t$. Evidentemente,

$$\mathbf{M}[\eta - \eta_\varepsilon]^2 = \int\limits_{|y|>\varepsilon} y^2 p(y)\,dy =$$

$$= \frac{\Delta t}{\sqrt{2\pi}\,\sigma(t, x)} \int\limits_{|z|>\frac{\varepsilon}{\sqrt{\Delta t}}} z^2 e^{-(z-a)^2/2\sigma(t, x)^2}\,dz = o(\Delta t)$$

uniformemente según t y x para $|a(t, x)| \leqslant C$, $C_1 \leqslant \sigma(t, x) \leqslant C_2$. Aquí nos interesa la magnitud $\Delta\xi(t)$, para la cual

$$\mathbf{M}[\Delta\xi(t) - \eta]^2 = o(\Delta t).$$

De la desigualdad $\|\Delta\xi(t) - \eta_\varepsilon\| \leqslant \|\Delta\xi - \eta\| + \|\eta - \eta_\varepsilon\|$ obtenemos también que

$$\mathbf{M}[\Delta\xi(t) - \eta_\varepsilon]^2 = o(\Delta t).$$

Ya que $|\Delta\xi(t)|^2 \leqslant 2[|\eta_\varepsilon|^2 + |\Delta\xi(t) - \eta_\varepsilon|^2]$, entonces

$$\chi_A |\Delta\xi(t)|^2 \leqslant 2[\chi_A \varepsilon^2 + |\Delta\xi(t) - \eta_\varepsilon|^2],$$

donde $\chi_A = 1$, al aparecer el suceso A, en caso contrario $\chi_A = 0$. Para el suceso $A = \{|\Delta\xi(t)| > 2\varepsilon\}$ que entra en el suceso $B = \{|\Delta\xi(t) - \eta_\varepsilon| > \varepsilon\}$ tenemos

$$\mathbf{M}[\chi_A\,\Delta\xi(t)]^2 \leqslant 2[\varepsilon^2 \mathbf{P}(A) + \mathbf{M}[\Delta\xi(t) - \eta_\varepsilon]^2]$$

y

$$\mathbf{P}(A) \leqslant \mathbf{P}(B) \leqslant \frac{1}{\varepsilon^2}\mathbf{M}[\Delta\xi(t) - \eta_\varepsilon]^2.$$

Como resultado obtenemos

$$\mathbf{M}[\chi_A\,\Delta\xi(t)]^2 = \mathbf{M}[\Delta\xi(t) - (\Delta\xi(t))_{2\varepsilon}]^2 \leqslant$$

$$\leqslant 4\mathbf{M}[\Delta\xi(t) - \eta_\varepsilon]^2 = o(\Delta t)$$

para cualquier $\varepsilon > 0$. Ahora ya es fácil deducir que

$$\mathbf{P}\{|\Delta\xi(t)| > 2\varepsilon\} \leqslant \frac{1}{4\varepsilon^2}\mathbf{M}[\Delta\xi(t) - (\Delta\xi(t))_{2\varepsilon}]^2 = o(\Delta t),$$

$$|\mathbf{M}\Delta\xi(t) - \mathbf{M}(\Delta\xi(t))_{2\varepsilon}| \leqslant \frac{1}{2\varepsilon}\mathbf{M}[\Delta\xi(t) - (\Delta\xi(t))_{2\varepsilon}]^2 = o(\Delta t)$$

y, ya que $\| \Delta\xi(t) \|^2 = \sigma^2(t, x)\,\Delta t + o(\Delta t)$,

$$|\mathbf{M}(\Delta\xi(t))^2 - \mathbf{M}(\Delta\xi(t))^2_{2\varepsilon}| = \| \Delta\xi(t) \|^2 - \| (\Delta\xi(t))_{2\varepsilon} \|^2 \leqslant$$
$$\leqslant 2 \| \Delta\xi(t) \| \cdot \| (\Delta\xi(t)) - (\Delta\xi(t))_{2\varepsilon} \| = o(\Delta t).$$

de donde, si sustituimos 2ε por ε, se deducen directamente las condiciones (8.27).

2. Ecuaciones diferenciales de Kolmogórov. Examinemos el proceso aleatorio de Márkov $\xi(t)$. Consideraremos que para cualesquiera $s \leqslant t$ con la condición de que $\xi(s) = x$, la magnitud aleatoria $\xi(t)$ tiene la densidad condicional de probabilidad $p(s, x, t, y)$, $-\infty < y < \infty$, de modo que en el tiempo desde s hasta t, el proceso pasa del estado x a un estado determinado y, $y' \leqslant y \leqslant y''$, con la probabilidad

$$\mathbf{P}\{y' \leqslant \xi(t) \leqslant y'' \mid \xi(s) = x\} = \int_{y'}^{y''} p(s, x, t, y)\,dy.$$

De hecho $p(s, x, t, y)$ nos da la densidad condicional de probabilidad de la magnitud $\xi(t)$ para cualquier condición sobre la magnitud $\xi(u)$, $u \leqslant s$, incluyendo en sí $\xi(s) = x$, ya que para el proceso aleatorio de Márkov, la magnitud $\xi(t)$ no depende de $\xi(u)$, $u \leqslant s$, para el estado conocido $\xi(s) = x$.

Demostremos que la densidad de paso de la probabilidad $p(s, x, t, y)$ satisfase la llamada *ecuación de Kolmogórov-Chapman*:

$$p(s, x, t, y) = \int_{-\infty}^{\infty} p(s, x, u, z)\,p(u, z, t, y)\,dz \qquad (8.28)$$

para cualesquiera $s \leqslant u \leqslant t$.

En efecto, para las condiciones $\xi(s) = x$, $\xi(u) = z$, la magnitud $\xi(t)$ no depende de x y tiene una densidad condicional de probabilidad, igual a $p(u, z, t, y)$; examinando la distribución de probabilidades de las magnitudes $\xi(u)$ y $\xi(t)$ para el valor fijado $\xi(s) = x$, hallamos que

$$p(s, x, u, z)\,p(u, z, t, y), \quad -\infty < z, y < \infty,$$

es la densidad conjunta de probabilidad de las magnitudes $\xi(u)$, $\xi(t)$[1]; integrando esta densidad conjunta por z, $-\infty < z < \infty$, obte-

[1]) Análogamente, para cualesquiera $s \leqslant t_1 \leqslant \ldots \leqslant t_n$, $p(s, x, t_1, y_1) \times p(t_1, y_1, t_2, y_2) \ldots p(t_{n-1}, y_{n-1}, t_n, y_n)$, $-\infty < y_1, \ldots, y_n < \infty$, es la densidad conjunta de probabilidad de las magnitudes aleatorias $\xi(t_1)$, $\xi(t_2)$, $\ldots$ $\ldots$, $\xi(t_n)$ para el valor fijado $\xi(s) = x$.

nemos la densidad de probabilidad $p(s, x, t, y)$ de la magnitud $\xi(t)$ (para el valor fijado $\xi(s) = x$).

Sea $\xi(t)$ un proceso de difusión. Evidentemente las condiciones (8.27) se pueden expresar directamente por su densidad de paso $p(s, x, t, y)$ del modo siguiente:

$$\begin{gathered}\int\limits_{|y-x|>\varepsilon} p(t, x, t+\Delta t, y)\, dy = o(\Delta t),\\ \int\limits_{|y-x|\leqslant\varepsilon} (y-x)\, p(t, x, t+\Delta t, y)\, dy = a(t, x)\,\Delta t + o(\Delta t), \qquad (8.29)\\ \int\limits_{|y-x|\leqslant\varepsilon} (y-x)^2\, p(t, x, t+\Delta t, y)\, dy = b(t, x)\,\Delta t + o(\Delta t),\end{gathered}$$

donde, recordamos una vez más que $o(\Delta t)$ significa una magnitud infinitamente pequeña de orden superior en relación a $\Delta t \to 0$, uniformemente según t en cada intervalo limitado $t_0 \leqslant t \leqslant t_1$ y $b(t, x) = \sigma^2(t, x)$.

Teorema 2. *Supongamos que la densidad de paso de la probabilidad $p(s, x, t, y)$ tiene las derivadas continuas $\frac{\partial p}{\partial s}$, $\frac{\partial p}{\partial x}$ y $\frac{\partial^2 p}{\partial x^2}$ uniformemente continuas según y en cada intervalo limitada $y' \leqslant y \leqslant y''$. Entonces ésta satisface a la ecuación diferencial*

$$-\frac{\partial p}{\partial s} = a(s, x)\frac{\partial p}{\partial x} + \frac{1}{2} b(s, x)\frac{\partial^2 p}{\partial x^2}. \qquad (8.30)$$

Demostración. Tomemos una función continua arbitraria $\varphi(x)$, que sea igual a cero fuera de un intervalo finita y supongamos que

$$\varphi(s, x) = \int\limits_{-\infty}^{\infty} \varphi(y)\, p(s, x, t, y)\, dy.$$

De la ecuación de Kolmogórov—Chapman se deduce que para cualesquiera $t_0 \leqslant s \leqslant u \leqslant t$

$$\varphi(s, x) = \int\limits_{-\infty}^{\infty} \varphi(y) \int\limits_{-\infty}^{\infty} p(s, x, u, z)\, p(u, z, t, y)\, dy =$$

$$= \int\limits_{-\infty}^{\infty} \varphi(u, z)\, p(s, x, u, z)\, dz.$$

Evidentemente, la función $\varphi(s, x)$ tiene las derivadas continuas $\frac{\partial \varphi}{\partial s}$, $\frac{\partial \varphi}{\partial x}$ y $\frac{\partial^2 \varphi}{\partial x^2}$. Desarrollemos la función $\varphi(u, z)$ en los alrededores del punto x (para el valor fijado u) según la fórmula de Taylor

$$\varphi(u, z) - \varphi(u, x) = \frac{\partial \varphi(u, x)}{\partial x}(z-x) + \frac{1}{2}\left[\frac{\partial^2 \varphi(u, x)}{\partial x^2} + O(\delta_\varepsilon)\right](z-x)^2,$$

donde

$$\delta_\varepsilon = \sup_{|z-x|\leqslant\varepsilon} \left| \frac{\partial^2 \varphi(u, z)}{\partial x^2} - \frac{\partial^2 \varphi(u, x)}{\partial x^2} \right| \to 0$$

para $\varepsilon \to 0$. De las relaciones (8.29) obtenemos

$$\varphi(s, x) - \varphi(u, x) = \int_{-\infty}^{\infty} [\varphi(u, z) - \varphi(u, x)]\, p(s, x, u, z)\, dz =$$

$$= \int_{|z-x|\leqslant\varepsilon} [\varphi(u, z) - \varphi(u, x)]\, p(s, x, u, z)\, dz +$$

$$+ o(u-s) = \frac{\partial \varphi(u, x)}{\partial x} \int_{|z-x|\leqslant\varepsilon} (z-x)\, p(s, x, u, z)\, dz +$$

$$+ \frac{1}{2}\left[\frac{\partial^2 \varphi(u, x)}{\partial x^2} + O(\delta_\varepsilon)\right] \int_{|z-x|\leqslant\varepsilon} (z-x)^2\, p(s, x, u, z)\, dz +$$

$$+ o(u-s) = \left\{ a \frac{\partial \varphi}{\partial x} + \frac{1}{2} b \left[\frac{\partial^2 \varphi}{\partial x^2} + O(\delta_\varepsilon)\right] \right\} (u-s) + o(u-s),$$

donde $O(\delta_\varepsilon) \to 0$ para $\varepsilon \to 0$, de donde se ve que

$$\lim_{u \to s} \frac{\varphi(s, u) - \varphi(u, x)}{u-s} = a(s, x) \frac{\partial \varphi}{\partial x} + \frac{1}{2} b(s, x) \frac{\partial^2 \varphi}{\partial x^2},$$

y, de tal modo,

$$-\frac{\partial \varphi}{\partial s} = a(s, x) \frac{\partial \varphi}{\partial x} + \frac{1}{2} b(s, x) \frac{\partial^2 \varphi}{\partial x^2}.$$

Teniendo en cuenta la definición de la función $\varphi(s, x)$, esta ecuación se puede escribir de la siguiente forma:

$$\int_{-\infty}^{\infty} \varphi(y) \left[\frac{\partial p}{\partial s} + a(s, x) \frac{\partial p}{\partial x} + \frac{1}{2} b(s, x) \frac{\partial^2 p}{\partial x^2} \right] dy = 0,$$

donde, recordemos $\varphi(y)$ es una función continua arbitraria, igual a cero fuera del intervalo finito y, por consiguiente, se deberá cumplir la igualdad

$$\frac{\partial p}{\partial s} + a(s, x) \frac{\partial p}{\partial x} + \frac{1}{2} b(s, x) \frac{\partial^2 p}{\partial x^2} = 0.$$

El teorema queda demostrado.

Señalemos que la densidad de paso de la probabilidad $p(s, x, t, y)$ coincide con la tal llamada solución fundamental de la ecuación (8.30), que se determina por la misma condición que para cualquier función continua limitada $\varphi(x)$

$$\lim_{t \to s} \int_{-\infty}^{\infty} \varphi(y)\, p(s, x, t, y)\, dy = \varphi(x).$$

Teorema 3. *Supongamos que se tienen las derivadas continuas*

$$\frac{\partial}{\partial t} p(s, x, t, y), \quad \frac{\partial}{\partial y}[a(t, y) p(s, x, t, y)], \quad \frac{\partial^2}{\partial y^2}[b(t, y) p(s, x, t, y)].$$

Entonces la densidad de paso $p(s, x, t, y)$ satisface a la ecuación diferencial

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial y}[a(t, y) p(s, x, t, y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(t, y) p(s, x, t, y)]. \quad (8.31)$$

Demostración. Exactamente igual que en la demostración del teorema 2, se puede obtener fácilmente, que para cualquier función $\varphi(x)$ dos veces diferenciable continuamente, que sea igual a cero fuera de un intervalo finito, existe el límite

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\int_{-\infty}^{\infty} \varphi(y) p(s, x, t+\Delta t, y)\, dy - \varphi(x)\right] =$$

$$= a(t, x)\varphi'(x) + \frac{1}{2} b(t, x)\varphi''(x).$$

Tenemos

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} p(s, x, t, y)\varphi(y)\, dy =$$

$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\int_{-\infty}^{\infty} p(s, x, t+\Delta t, y)\varphi(y)\, dy - \int_{-\infty}^{\infty} p(s, x, t, z)\varphi(z)\, dz\right] =$$

$$= \int_{-\infty}^{\infty} p(s, x, t, z) \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[\int_{-\infty}^{\infty} p(s, z, t+\Delta t, y)\varphi(y)\, dy - \varphi(z)\right] dz =$$

$$= \int_{-\infty}^{\infty} p(s, x, t, z)\left[a(t, z)\varphi'(z) + \frac{1}{2} b(t, z)\varphi''(z)\right] dz.$$

Integrando por partes la última expresión, obtenemos

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} p(s, x, t, y)\varphi(y)\, dy = \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial t} p(s, x, t, y)\right]\varphi(y)\, dy =$$

$$= \int_{-\infty}^{\infty}\left\{-\frac{\partial}{\partial y}[a(t, y) p(s, x, t, y)] + \frac{1}{2}\frac{\partial^2}{\partial y^2}[b(t, y) p(s, x, t, y)]\right\}\varphi(y)\, dy,$$

de donde, se deduce la igualdad (8.31) debido a la arbitrariedad de la función $\varphi(y)$. El teorema queda demostrado.

La ecuación (8.30) se llama *inversa*, y la ecuación (8.31) es *la ecuación directa de Kolmogórov*.

El método de las ecuaciones diferenciales descrito anteriormente es utilizable al examinar, no sólo, la densidad de probabilidad $p(s, x, t, y)$ de la magnitud $\xi(t)$, sino también para otras magnitudes aleatorias, ligadas con el comportamiento del proceso de difusión.

Examinemos, por ejemplo, el momento τ de la primera salida del proceso aleatorio $\xi(t)$, $t \geqslant t_0$ del intervalo (c_1, c_2) (consideramos, que la trayectoria $\xi = \xi(t)$ del proceso de difusión es una función continua y por eso la magnitud τ es el momento en que por primera vez esta trayectoria alcanza los valores $x = c_1$ o $x = c_2$).

Designemos por $\mathbf{P}_{s, x}(A)$ la probabilidad del suceso A y por $\mathbf{M}_{s, x}\eta$, la esperanza matemática de la magnitud aleatoria η en condiciones que $\xi(s) = x$. Según la suposición, para el proceso de difusión $\xi = \xi(t)$ tenemos

$$\mathbf{P}_{s, x}\{|\xi(s + \Delta s) - x| > \varepsilon\} = o(\Delta s) \tag{8.32}$$

uniformemente por s en cada intervalo finito (cualquiera que sea el número fijado $\varepsilon > 0$), de donde se deduce que para la condición $\xi(s) = x$, la probabilidad de encontrarse durante el intervalo de tiempo de s a $s + \Delta s$ en un determinado punto y, $|y - x| > \varepsilon$, es la magnitud $o(\Delta s)$ de un orden infinitesimal superior del mismo tipo que en la relación (8.32); por consiguiente, también

$$\mathbf{P}_{s, x}\{\tau \leqslant s + \Delta s\} = o(\Delta s) \tag{8.33}$$

uniformemente por s en cada intervalo finito.

Supongamos que para la condición $\xi(s) = x$ la magnitud aleatoria τ tiene la densidad de probabilidad $p_{s, x}(t)$. Examinemos la función

$$u(s, x) = \mathbf{M}_{s, x} \exp\left\{\int_s^\tau U(t)\,dt\right\} = \int_{-\infty}^{\infty} \exp\left\{\int_s^t U(u)\,du\right\} p_{s, x}(t)\,dt,$$

donde la función continua $U(t)$ tiene una parte real $\operatorname{Re} U(t) \leqslant 0$ de modo que $\exp\left\{\int_s^\tau U(t)\,dt\right\}$ es una magnitud limitada. Para $U(t) = i\lambda t$ tenemos, por ejemlo,

$$u(s, x) = \int_{-\infty}^{\infty} e^{i\lambda t} p_{s, x}(t)\,dt$$

(donde $p_{s, x}(t) = 0$ para $t < s$), es decir, $u(s, x)$ nos da la función característica de la magnitud τ en el punto λ, $-\infty < \lambda < \infty$.

Para cualquier magnitud limitada η, evidentemente,

$$\begin{aligned}\mathbf{M}_{s, x}\eta = {} & \mathbf{M}_{s, x}\{\eta \mid \tau > s + \Delta s\}\,\mathbf{P}\{\tau > s + \Delta s\} + \\ & + \mathbf{M}_{s, x}\{\eta \mid \tau \leqslant s + \Delta s\}\,\mathbf{P}\{\tau \leqslant s + \Delta s\} = \\ & = \mathbf{M}_{s, x}\{\eta \mid \tau > s + \Delta s\}\,\mathbf{P}\{\tau > s + \Delta s\} + o(\Delta s),\end{aligned}$$

donde $o(\Delta s)$ es una magnitud del mismo tipo que en la relación (8.33). Tenemos

$$\mathbf{M}_{s,x}\left\{\exp\int_s^\tau U(t)\,dt \mid \tau > s+\Delta s\right\} =$$

$$= \mathbf{M}_{s,x}\left\{\mathbf{M}\left[\exp\int_s^\tau U(t)\,dt \mid \tau > s+\Delta s,\ \xi(s+\Delta s)=\eta\right] \mid \tau > s+\Delta s\right\} =$$

$$= \mathbf{M}_{s,x}\left\{\exp\left[\int_s^{s+\Delta s} U(t)\,dt\right] \mathbf{M}_{s+\Delta s,\eta} \exp\left[\int_{s+\Delta s}^\tau U(t)\,dt\right]\right\},$$

ya que las relaciones $\tau > s+\Delta s$, $\xi(s+\Delta s)=\eta$ representan en sí las condiciones sobre el comportamiento de la trayectoria $\xi(t)$, $t \leqslant s+\Delta s$, y para el proceso de Márkov $\xi = \xi(t)$ pueden ser sustituidas (como fue hecho anteriormente) sencillamente por la condición $\xi(s+\Delta s)=\eta$. Teniendo en cuenta que

$$\exp\left[\int_s^{s+\Delta s} U(t)\,dt\right] = 1 + U(s)\,\Delta s + o(\Delta s) = \frac{1}{1-U(s)\,\Delta s} + o(\Delta s),$$

obtenemos

$$u(s,x) = \mathbf{M}_{s,x}\left\{\frac{1}{1-U(s)\,\Delta s}\,u[s+\Delta s,\ \xi(s+\Delta s)]\right\} + o(\Delta s),$$

o bien,

$$\int_{-\infty}^{\infty} [u(s+\Delta s,\ y) - u(s,x)]\,p(s,x,s+\Delta s,\ y)\,dy + \\ + U(s)\,u(s,x)\,\Delta s + o(\Delta s) = 0,$$

lo pue para el proceso de difusión es equivalente a la relación

$$\int_{|y-x|\leqslant\varepsilon} [u(s+\Delta s,\ y) - u(s,x)]\,p(s,x,s+\Delta s,\ y)\,dy + \\ + U(s)\,u(s,x)\,\Delta s + o(\Delta s) = 0.$$

Supongamos ahora, que la función $u(s,x)$ tiene las derivadas continuas $\frac{\partial u}{\partial s}$, $\frac{\partial u}{\partial x}$ y $\frac{\partial^2 u}{\partial x^2}$. Entonces, igualmente que al deducir la ecuación diferencial (8.30), obtenemos

$$\int_{|y-x|\leqslant\varepsilon} \left[\frac{\partial u(s,x)}{\partial s}\Delta s + \frac{\partial u(s,x)}{\partial x}\Delta x + \frac{1}{2}\frac{\partial^2 u(s,x)}{\partial x^2}(1+O(\delta))\,\Delta x^2\right] \times \\ \times p(s,x,s+\Delta s,\ y)\,dy + U(s)\,u(s,x)\,\Delta s + o(\Delta s) =$$

$$= \left[\frac{\partial u(s, x)}{\partial s} + a(s, x)\frac{\partial u(s, x)}{\partial x} + \right.$$

$$\left. + \frac{1}{2} b(s, x)\frac{\partial^2 u(s, x)}{\partial x^2}(1 + O(\delta)) + U(s)\, u(s, x)\right]\Delta s + o(\Delta s) = 0,$$

donde $O(\delta) \to 0$ para $\varepsilon \to 0$, de donde se ve que la función $u(s, x)$ satisface a la ecuación diferencial

$$\frac{\partial u}{\partial s} + a\frac{\partial u}{\partial x} + \frac{1}{2} b\frac{\partial^2 u}{\partial x^2} + Uu = 0. \tag{8.34}$$

Además de esto, es evidente que para $x = c_1$ y $x = c_2$ (cuando $\tau = s$), según la definición de la función $u(s, x)$ tenemos

$$u(s, c_1) = u(s, c_2) \equiv 1.$$

CAPITULO IV

ALGUNAS TAREAS DE PRONOSTICACIÓN, FILTRADO. Y REGULACIÓN DE LOS PROCESOS ALEATORIOS

§ 1. TAREA GENERAL SOBRE LA APROXIMACION OPTIMA. EJEMPLOS

La mayoría de las tareas de pronosticación, filtrado y regulación de los procesos aleatorios descansan sobre la base del problema fundamental de hallar una valuación suficientemente buena de una determinada magnitud desconocida ξ según los valores existentes $\eta(t)$ de un determinado proceso aleatorio, en tal o cual intervalo de tiempo $a \leqslant t \leqslant b$; con ello corrientemente $\xi = \xi(t)$ es el valor en el momento variable de tiempo t (o relacionado al futuro) de otro proceso aleatorio, ligado de alguna forma con el proceso «observado» $\eta(t)$, $a \leqslant t \leqslant b$.

El problema más sencillo de tal género fue examinado por nosotros ya en el § 4 del cap. I, cuando tratamos la aproximación óptima de la magnitud ξ por combinaciones lineales de la forma $\sum_{k=1}^{n} c_k \eta_k$, donde η_k, $k = 1, \ldots, n$ son determinadas magnitudes dadas. Precisamente se exigía hallar la magnitud $\hat{\xi} = \sum_{k=1}^{n} c_k^0 \eta_k$ de tal modo que

$$\| \xi - \hat{\xi} \| = \min_{c_k} \Big\| \xi - \sum_{k=1}^{n} c_k \eta_k \Big\|. \tag{1.1}$$

Como se indicó, si introducimos el espacio lineal H de $(n+1)$ dimensiones de todas las magnitudes $\eta = \sum_{k=0}^{n} c_k \eta_k$ (donde $\eta_0 = \xi$ y $c_0, \ldots, c_n$ son coeficientes reales arbitrarios) con el producto escalar

$$\langle \eta_1, \eta_2 \rangle = \mathbf{M}(\eta_1 \cdot \eta_2) \tag{1.2}$$

y con la distancia correspondiente

$$\| \eta_1 - \eta_2 \| = \sqrt{\mathbf{M}(\eta_1 - \eta_2)^2}, \tag{1.3}$$

entonces la magnitud $\hat{\xi} = \sum_{k=1}^{n} c_k^0 \eta_k$ que satisface a la condición (1.1) significa geométricamente la base de la perpendicular bajada desde el punto $\xi \in H$ al subespacio L de todas las magnitudes $\eta = \sum_{k=1}^{n} c_k \eta_k$ y se determina, unívocamente, por la condición de ortogonalidad de la diferencia $\xi - \hat{\xi}$ hacia el subespacio L

$$\langle \xi - \hat{\xi}, \eta \rangle = 0, \quad \eta \in L, \tag{1.4}$$

esto es equivalente al siguiente sistema lineal de ecuaciones con relación a $c_1^0, \ldots, c_n^0$

$$\sum_{k=1}^{n} c_k^0 \langle \eta_k, \eta_j \rangle = \langle \xi, \eta_j \rangle, \quad j = 1, \ldots, n. \tag{1.5}$$

Examinemos la tarea general sobre la aproximación de una magnitud ξ por las magnitudes η de un conjunto L determinado, que es como se dice, *el hiperplano*, y que significa lo siguiente: para cualquier elemento $\eta_0 \in L$ el conjunto de magnitudes $\Delta = \eta - \eta_0$, $\eta \in L$ forma *un espacio lineal*, o sea, contiene junto con cualquier Δ_1, Δ_2 también su combinación lineal $c_1\Delta_1 + c_2\Delta_2$ (donde c_1, c_2 son coeficientes reales arbitrarios).

La magnitud $\hat{\xi} \in L$ la denominamos *aproximación óptima* de ξ, si

$$\| \xi - \hat{\xi} \| = \min_{\eta \in L} \| \xi - \eta \|. \tag{1.6}$$

Lema sobre la perpendicular. *La condición* (1.6) *es equivalente a lo siguiente*:

$$\langle \xi - \hat{\xi}, \eta - \hat{\xi} \rangle = 0 \text{ para todos los } \eta \in L. \tag{1.7}$$

Antes de demostrar esta afirmación, demos a la relación (1.7) una interpretación geométrica sencilla. Si designamos, precisamente, por $\hat{L}$ el espacio lineal de todos los elementos $\eta - \hat{\xi}$, $\eta \in L$, entonces la relación (1.7) significará que el producto escalar de los elementos $\xi - \hat{\xi}$ y $\Delta = \eta - \hat{\xi}$ será igual a cero para todos los $\Delta \in L$, en otras palabras, la diferencia $\xi - \hat{\xi}$ es perpendicular al subespacio $\hat{L}$ (fig. 24).

La magnitud $\hat{\xi}$ se llama *proyección* de la magnitud ξ en el hiperplano L y la diferencia $\xi - \hat{\xi}$, perpendicular (desde el punto ξ).

Señalemos también que en el caso cuando el hiperplano L por sí mismo es un espacio lineal ($L = \hat{L}$), la condición (1.7) es equivalente a la condición (1.4).

Demostración del lema. Supongamos que $\hat{\xi} \in L$ es la aproximación óptima de ξ. Evidentemente,

$$\| \xi - \hat{\xi} \| = \min \| \xi - \hat{\xi} - \Delta \|,$$

donde el mínimo se toma por todas las diferencias $\Delta = \eta - \hat{\xi}$, $\eta \in L$. Pero el conjunto de tales diferencias forma el espacio lineal $\hat{L}$ y en particular, en $\hat{L}$ se contiene junto con el elemento Δ el elemento $\lambda\Delta$, donde λ es un número real. Fijando cualquier valor $\Delta \neq 0$ y suponiendo que

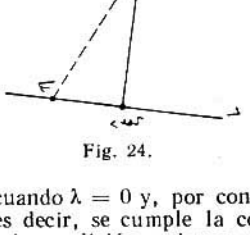

Fig. 24.

$$A^2 = \| \xi - \hat{\xi} \|^2, \; B = \langle \xi - \hat{\xi}, \Delta \rangle, \; C = \| \Delta \|^2,$$

tendremos

$$\min_{\lambda} \| \xi - \hat{\xi} - \lambda\Delta \|^2 = \min_{\lambda} \times \times (A^2 + 2B\lambda + \lambda^2 C^2) = A^2.$$

Se ve que el mínimo de la forma cuadrática $A^2 + 2B\lambda + \lambda C^2$ se alcanza cuando $\lambda = 0$ y, por consiguiente, el coeficiente B es igual a cero, es decir, se cumple la condición (1.7). A su vez, si está cumplida esta condición, entonces para cualquier magnitud $\eta \in L$ se tiene

$$\| \xi - \eta \|^2 = \| (\xi - \hat{\xi}) - (\eta - \hat{\xi}) \|^2 = \| \xi - \hat{\xi} \|^2 + 2 \langle \xi - \hat{\xi}, \eta - \hat{\xi} \rangle + \| \eta - \hat{\xi} \|^2 = \| \xi - \hat{\xi} \|^2 + \| \eta - \hat{\xi} \|^2 \geqslant \| \xi - \hat{\xi} \|^2,$$

de donde se ve, que $\hat{\xi}$ es la aproximación óptima de la magnitud ξ. El lema queda demostrado.

Señalemos, que *sólo existe la magnitud única* $\hat{\xi} \in L$, *que satisface la condición* (1.7) (*es decir, también la condición* (1.6)).

En efecto, si $\hat{\xi}_1$, $\hat{\xi}_2$ satisfacen las relaciones

$$\langle \xi - \hat{\xi}_1, \eta - \hat{\xi}_1 \rangle = 0, \quad \langle \xi - \hat{\xi}_2, \eta - \hat{\xi}_2 \rangle = 0$$

para todos los $\eta \in L$, entonces, en particular

$$\langle \xi - \hat{\xi}_1, \hat{\xi}_2 - \hat{\xi}_1 \rangle = 0, \; \langle \xi - \hat{\xi}_2, \hat{\xi}_2 - \hat{\xi}_1 \rangle = 0,$$

y tomando la diferencia de estas expresiones, obtenemos

$$\| \hat{\xi}_2 - \hat{\xi}_1 \| = 0, \text{ es decir, } \hat{\xi}_2 = \hat{\xi}_1.$$

Anteriormente hemos recordado como hallar la aproximación óptima (lineal) de la magnitud ξ con las combinaciones lineales $\sum_{k=1}^{n} c_k \eta_k$ de ciertas magnitudes dadas, η_k, $k = 1, \ldots, n$ (véase (1.1) — (1.5)).

¿Existe la aproximación óptima absoluta $\hat{\xi} = \varphi_0(\eta_1, \ldots, \eta_n)$ que satisfaga la condición (1.6) en la cual se toma el mínimo por todas las magnitudes posibles $\eta = \varphi(\eta_1, \ldots, \eta_n)$, que son funciones de los valores dados $\eta_1, \ldots, \eta_n$?

Teorema. *La aproximación óptima absoluta de la magnitud* ξ *por las magnitudes de la forma* $\eta = \varphi(\eta_1, \ldots, \eta_n)$ *se da por la fórmula*

$$\hat{\xi} = \mathbf{M}(\xi \mid \eta_1, \ldots, \eta_n), \tag{1.8}$$

donde $\mathbf{M}(\xi \mid \eta_1, \ldots, \eta_n)$ *significa la esperanza matemática condicional de la magnitud* ξ *para los valores fijados* $\eta_1, \ldots, \eta_n$.

Demostración. Evidentemente, el conjunto L de todas las magnitudes de la forma $\eta = \varphi(\eta_1, \ldots, \eta_n)$ para los cuales $\| \eta \|^2 = \mathbf{M}\eta^2 < \infty$ es un espacio lineal. Ya que $\mathbf{M}\hat{\xi}^2 < \infty$, entonces a este espacio le pertenece también la magnitud $\hat{\xi}$ indicada en (1.8), que es una función determinada de los valores $\eta_1, \ldots, \eta_n$:

$$\hat{\xi} = \varphi_0(\eta_1, \ldots, \eta_n) \in L,$$

$$\hat{\xi}^2 = [\mathbf{M}(\xi \mid \eta_1, \ldots \eta_n)]^2 \leqslant \mathbf{M}(\xi^2 \mid \eta_1, \ldots, \eta_n),$$

y

$$\mathbf{M}\hat{\xi}^2 \leqslant \mathbf{M}[\mathbf{M}(\xi^2 \mid \eta_1, \ldots, \eta_n)] = \mathbf{M}\xi^2 < \infty.$$

Para la demostración de que la magnitud $\hat{\xi} = \varphi_0(\eta_1, \ldots, \eta_n)$ da la aproximación óptima absoluta, es suficiente comprobar la condición correspondiente (1.7) que en nuestro caso significa que

$$\mathbf{M}[(\xi - \hat{\xi}) \cdot \varphi(\eta_1, \ldots, \eta_n)] = 0$$

para cualquier magnitud $\varphi(\eta_1, \ldots, \eta_n) \in L$. Sin embargo, para los valores fijados $\eta_1, \ldots, \eta_n$

$$\mathbf{M}[(\xi - \hat{\xi}) \cdot \varphi(\eta_1, \ldots, \eta_n) \mid \eta_1, \ldots, \eta_n] = \varphi(\eta_1, \ldots, \eta_n)[\mathbf{M}(\xi \mid \eta_1, \ldots, \eta_n) - \hat{\xi}] = 0,$$

y por consiguiente,

$$\mathbf{M}\{(\xi - \hat{\xi}) \cdot \varphi(\eta_1, \ldots, \eta_n)\} = \mathbf{M}\{\mathbf{M}[(\xi - \hat{\xi}) \cdot \varphi(\eta_1, \ldots, \eta_n) \mid \eta_1, \ldots, \eta_n]\} = 0.$$

El teorema queda demostrado.

Ejemplo (*tarea sobre un complejo de aparatos*). Imaginémonos que existen n aparatos destinados a la medición de una magnitud desconocida a, con ello cada aparato admite (en dependencia del caso) un error en las mediciones indicando correspondientemente $X_k = a + \Delta_k$ en lugar del valor verdadero a ($k = 0, \ldots, n - 1$). Surge la tarea de «unir» estos n aparatos en un complejo de tal forma que las indicaciones del «aparato en complejo» $X = a + \Delta$ den el error mínimo: $\| \Delta \| = \text{mín}$.

Está claro, que durante la elaboración de los errores de medición $\Delta_1, \ldots, \Delta_n$ al principio es necesario excluir la propia magnitud a desconocida de los datos existentes $X_1, \ldots, X_n$. Para esto se puede pasar a las diferencias

$$X_k - X_0 = \Delta_k - \Delta_0 = \xi_k, \; k = 1, \ldots, n-1,$$

intentando luego hallar la aproximación óptima para el error desconocido $\Delta_0 (= \xi_0)$ por las *magnitudes conocidas* $\xi_1, \ldots, \xi_{n-1}$.

Teniendo tal aproximación óptima

$$\hat{\Delta}_0 = \varphi_0(\xi_1, \ldots, \xi_{n-1}) = \varphi_0(X_1 - X_0, \ldots, X_{n-1} - X_0),$$

se puede realizar «el aparato en complejo» del siguiente modo:

$$X = X_0 - \varphi_0(X_1 - X_0, \ldots, X_{n-1} - X_0)$$

en el cual el error $\Delta = X - a$ será

$$\Delta = \Delta_0 - \varphi_0(\xi_1, \ldots, \xi_{n-1}).$$

Es natural considerar que los errores $\Delta_0, \Delta_1, \ldots, \Delta_{n-1}$ de los distintos aparatos son magnitudes aleatorias i n d e p e n d i e n t e s. Supondremos que éstas tienen una misma distribución de probabilidades (con el valor medio nulo). Entonces la matriz de correlación $\{R_{kj}\}$ de las magnitudes $\xi_0, \xi_1, \ldots, \xi_{n-1}$ tiene la forma siguiente:

$$R_{00} = M\Delta_0^2 = \sigma^2, \quad R_{k0} = -\sigma^2, \quad R_{kk} = 2\sigma^2,$$
$$k = 1, \ldots, n-1.$$
$$R_{kj} = \sigma^2, \; k \neq j; \; k, j = 1, \ldots, n-1.$$

De las ecuaciones (1.5) obtenemos que la aproximación lineal óptima es

$$-\frac{1}{n}\sum_{k=1}^{n-1}\xi_k = \frac{1}{n}\sum_{k=1}^{n-1}[X_0 - X_k],$$

y el error correspondiente del «aparato en complejo» será

$$\Delta = \Delta_0 - \frac{1}{n}\sum_{k=1}^{n-1}[\Delta_0 - \Delta_k] = \frac{1}{n}\sum_{k=1}^{n}\Delta_k.$$

El "aparato en complejo" se realiza del modo siguiente:

$$X = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Supongamos ahora que el error de cada aparato por separado tiene *una distribución uniforme de probabilidades* (en un segmento $[-d, d]$).

Hallemos la aproximación óptima absoluta

$$\varphi_0(\xi_1, \ldots, \xi_{n-1}) = \mathbf{M}(\Delta_0 \mid \xi_1, \ldots, \xi_{n-1}).$$

La densidad conjunta de probabilidad para las magnitudes iniciales $\Delta_0, \ldots, \Delta_{n-1}$ es

$$p_{\Delta_0, \ldots, \Delta_{n-1}}(y_0, \ldots, y_{n-1}) = \begin{cases} \frac{1}{(2d)^n} \text{ para } -d \leqslant y_k \leqslant d, \ k=0, \ldots, n-1, \\ 0 \quad \text{en caso contrario.} \end{cases}$$

Por consiguiente, la densidad conjunta de probabilidad para las magnitudes $\xi_0, \ldots, \xi_{n-1}$, ligadas con $\Delta_0, \ldots, \Delta_{n-1}$ por la transformación $\xi_0 = \Delta_0$, $\xi_k = \Delta_k - \Delta_0$, $k = 1, \ldots, n-1$, será

$$p_{\xi_0, \ldots, \xi_{n-1}}(x_0, \ldots, x_{n-1}) = \begin{cases} \frac{1}{(2d)^n} \text{ para } \alpha \leqslant x_0 \leqslant \beta, \\ 0 \quad \text{en caso contrario,} \end{cases}$$

donde

$$\alpha = \max(-d, -d - x_1, \ldots, -d, -x_{n-1}),$$
$$\beta = \min(d, d - x_1, \ldots, d - x_{n-1}),$$

ya que el jacobiano de la transformación indicada es igual a 1, y la desigualdad $\alpha \leqslant x_0 \leqslant \beta$ significa exactamente que los valores correspondientes $y_0, \ldots, y_{n-1}$ se encuentran entre los límites $-d \leqslant y_k \leqslant d$, $k = 0, \ldots, n-1$. Luego, la densidad de probabilidad de las magnitudes $\xi_1, \ldots, \xi_{n-1}$ es

$$p_{\xi_1, \ldots, \xi_{n-1}}(x_1, \ldots, x_{n-1}) = \int_{-\infty}^{\infty} p_{\xi_0 \xi_1, \ldots \xi_{n-1}}(x_0, x_1, \ldots, x_{n-1})\, dx_0 =$$

$$= \begin{cases} \frac{\beta - \alpha}{(2d)^n} \text{ para } \alpha \leqslant \beta, \\ 0 \quad \text{para } \alpha > \beta, \end{cases}$$

de modo que para $\tilde{\xi}$ la densidad condicional de probabilidad $p_{\xi_0}(x_0 \mid x_1, \ldots, x_{n-1})$ obtenemos la expresión

$$p_{\xi_0}(x_0 \mid x_1, \ldots, x_{n-1}) = \begin{cases} \frac{1}{\beta - \alpha} \text{ para } \alpha \leqslant \beta \\ 0 \quad \text{para } \alpha > \beta. \end{cases}$$

La función buscada $\varphi_0(x_1, \ldots, x_{n-1})$ es

$$\varphi_0(x_1, \ldots, x_{n-1}) = \int_{-\infty}^{\infty} x_0 p_{\xi_0}(x_0 \mid x_1, \ldots, x_{n-1})\, dx_0 = \frac{\alpha + \beta}{2}.$$

No es difícil ver que el «aparato en complejo» óptimo se realiza del modo siguiente:

$$X = \frac{X_* + X^*}{2},$$

donde X_* y X^* significan los valores mínimo y máximo en las indicaciones de diferentes aparatos $X_0, \ldots, X_{n-1}$ (*X es la semisuma de los valores extremos*).

En efecto, como se ve fácilmente,

$$\alpha = -d - (X_* - X_0), \quad \beta = d - (X^* - X_0)$$

y

$$X = X_0 - \varphi_0(\xi_1, \ldots, \xi_{n-1}) = X_0 - \frac{\alpha+\beta}{2} = \frac{X_*+X^*}{2},$$

donde

$$X_* = \text{mín}\,(X_0, \ldots, X_{n-1}), \quad X^* = \text{máx}\,(X_0, \ldots, X_{n-1}).$$

Es interesante comparar los errores correspondientes

$$\delta_0 = \left\| \frac{1}{n} \sum_{k=0}^{n-1} \Delta_k \right\| \quad \text{y} \quad \delta = \left\| \frac{1}{2}(\Delta_* + \Delta^*) \right\|$$

para las aproximaciones óptimas lineales y absolutas; aquí

$$\Delta_* = \text{mín}\,(\Delta_1, \ldots, \Delta_n), \quad \Delta^* = \text{máx}\,(\Delta_1, \ldots, \Delta_n).$$

Suponiendo para concretar que $d = 1/2$, tenemos

$$\delta_0^2 = \frac{1}{n} \int_{-1/2}^{1/2} x^2\, dx = \frac{1}{12n}$$

Luego, ya que

$$\Delta_* = \Delta_i, \ \Delta^* = \Delta_j \text{ para } \Delta_i \leqslant \Delta_k \leqslant \Delta_j \quad (k = 0, \ldots n-1),$$

entonces

$$\delta^2 = \frac{1}{4}\,\mathbf{M}\,(\Delta_* + \Delta^*)^2 =$$

$$= \frac{1}{4} \sum_{i,\,j\,(i \neq j)} \int_{-1/2}^{1/2} \int_{y_i}^{1/2} (y_i + y_j)^2 \underbrace{\int_{y_i}^{y_j} \cdots \int_{y_i}^{y_j}}_{n-2 \text{ veces}} dy_0 \ldots dy_n =$$

$$= \frac{1}{4}\, n(n-1) \int_{-1/2}^{1/2} \int_{y_0}^{1/2} \int_{y_0}^{y_1} \cdots \int_{y_0}^{y_1} (y_0 + y_1)^2\, dy_0 \ldots dy_{n-1} = \frac{1}{2(n+1)(n+2)}$$

Vemos que para valores grandes de n la aproximación óptima absoluta en nuestro caso tiene más preponderancia que la aproximación óptima lineal.

§. 2 PRONOSTICACION Y FILTRADO DE LOS PROCESOS ALEATORIOS ESTACIONARIOS

1. Tarea de pronosticación lineal. Sea $\xi(t)$ un proceso aleatorio estacionario real en un ámplio sentido que admite su representación espectral

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda) \tag{2.1}$$

con la densidad espectral $f(\lambda)$. Designemos por $H_{(-\infty,\, t)}$ el conjunto de todas las magnitudes

$$\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda), \tag{2.2}$$

donde $\varphi(\lambda)$ entra en el espacio correspondiente $L_{(-\infty,\, t)}(f)$ de todas las funciones $\varphi(\lambda)$, $\int_{-\infty}^{\infty} |\varphi(\lambda)|^2 f(\lambda)\, d\lambda$, que bien tienen la forma

$$\varphi(\lambda) = \sum_{k=1}^{n} c_k e^{i\lambda t_k}$$

(donde $c_1, \ldots, c_n$ son coeficientes reales y $t_1, \ldots, t_n \leqslant t$), o bien son los límites de las funciones de tal forma (en otras palabras, $H_{(-\infty,\, t)}$ es el espacio de las magnitudes, que bien son combinaciones lineales $\eta = \sum_{k=1}^{n} c_k \xi(t_k)$ de los valores del proceso estacionario examinado en los momentos $t_1 \ldots t_n \leqslant t$, o bien son los límites en la media cuadrática de tales combinaciones lineales — véase el punto 2 del § 7).

La tarea de pronosticación lineal consiste en hallar la aproximación óptima del valor «futuro» $\xi(t+\tau)$, $\tau > 0$, según las magnitudes $\eta \in H_{(-\infty,\, t)}$.

Demos la solución de esta tarea en el caso en que la densidad espectral es una función racional de λ y es representable en la forma

$$f(\lambda) = \frac{|P(i\lambda)|^2}{|Q(i\lambda)|^2}, \quad -\infty < \lambda < \infty, \tag{2.3}$$

(forma general de la densidad espectral racional) donde los polinomios $P(z)$ y $Q(z)$ tienen coeficientes reales, todas sus raíces se hallan en el semiplano izquierdo $\operatorname{Re} z < 0$ de la variable compleja z y el

polinomio $Q(z)$ no tiene raíces múltiples:

$$Q(z) = q_0 (z - q_1) \ldots (z - q_n). \tag{2.4}$$

Supongamos de momento, que $t = 0$. Abreviemos nuestras designaciones poniendo $H_0 = H_{(-\infty, 0)}$ y $L_0 = L_{(-\infty, 0)}(f)$. Buscaremos la función $\varphi_0(i\lambda) \in L_0$, con ayuda de la cual la aproximación óptima $\hat{\xi}$ para el valor futuro $\xi(\tau)$ puede ser dada como

$$\hat{\xi} = \int_{-\infty}^{\infty} \varphi_0(i\lambda)\, d\Phi(\lambda). \tag{2.5}$$

La condición general (1.7)

$$\langle \xi(\tau) - \hat{\xi},\ \eta \rangle = 0, \quad \eta \in L, \tag{2.6}$$

que determina unívocamente la aproximación óptima $\hat{\xi}$, en nuestro caso es equivalente a la condición

$$\langle \xi(\tau) - \hat{\xi},\ \xi(t) \rangle = 0 \quad \text{para } t \leqslant 0 \tag{2.7}$$

ya que de la igualdad (2.7) se deduce la igualdad (2.6) para todas las combinaciones lineales posibles $\eta = \sum_k c_k \xi(t_k)$, de las cuales se puede pasar a las magnitudes arbitrarias $\eta \in L$ por un paso límite.

La condición significa analíticamente (2.7) que

$$\int_{-\infty}^{\infty} \bar{e}^{i\lambda t} [e^{i\lambda\tau} - \varphi_0(i\lambda)]\, f(\lambda)\, d\lambda = 0 \quad \text{para } t \leqslant 0, \tag{2.8}$$

es decir, que la *transformación de Fourier de la función*

$$\psi(\lambda) = [e^{i\lambda\tau} - \varphi_0(i\lambda)]\, f(\lambda)] \tag{2.9}$$

se convierte en cero en el semieje negativo $t \leqslant 0$.

Con objeto de hallar la función $\varphi_0(i\lambda)$ de la condición (2.8) nos hacen falta las siguientes proposiciones.

Lema 1. *Si la función* $\psi(\lambda)$ *es analítica en el semiplano superior de la variable compleja* $\operatorname{Im} \lambda \geqslant 0$ *y decrece allí para* $\lambda \to \infty$ *de modo que* $|\psi(\lambda)| \leqslant C |\lambda|^{-1-\varepsilon}$, *donde* $\varepsilon > 0$, *entonces*

$$\int_{-\infty}^{\infty} \bar{e}^{i\lambda t} \psi(\lambda)\, d\lambda = 0 \ \text{para } t \leqslant 0 \tag{2.10}$$

Demostración. Para la función analítica $e^{i\lambda t}\psi(\lambda)$, $t \leqslant 0$, en el semiplano superior, la integral $\oint e^{i\lambda t}\psi(\lambda)\, d\lambda$ por el contorno limitado con el segmento $[-R, R]$ del eje real y con la semicircunferencia Γ de radio R (fig. 25), según el teorema de Cauchy es igual a cero.

Evidentemente, para $R \to \infty$

$$\left| \oint_{\Gamma} e^{i\lambda t} \psi(\lambda)\, d\lambda \right| \leqslant \max_{|\lambda|=R} |\psi(\lambda)|\, \pi R \leqslant C\, |R|^{-\varepsilon} \to 0,$$

y por esto para cualquier $t \leqslant 0$

$$\int_{-\infty}^{\infty} e^{i\lambda t} \psi(\lambda)\, d\lambda = \lim_{R\to\infty} \int_{-R}^{R} e^{i\lambda t} \psi(\lambda)\, d\lambda = \lim_{R\to\infty} \oint e^{i\lambda t} \psi(\lambda)\, d\lambda = 0,$$

que es lo que se exigía demostrar.

Lema 2. *La función racional*

$$\varphi_0(i\lambda) = \frac{\sum_{k=0}^{n-1} a_k (i\lambda)^k}{P(i\lambda)}, \tag{2.11}$$

en la que todas las raíces del denominador, o sea, del polinomio $P(z)$, de grado m, descansan en el semiplano izquierdo de la variable compleja $\mathrm{Re} z < 0$, *representable en la forma*

$$\varphi_0(i\lambda) = \sum_{k=0}^{n-m-1} c_k (i\lambda)^k + \int_0^{\infty} e^{-i\lambda t} c(t)\, dt, \tag{2.12}$$

donde $c(t)$ es una función integrable.

Demostración. Designemos por $\sum_{k=0}^{n-m-1} c_k (i\lambda)^k$ polinomio que se obtiene de la división de $\sum_{k=0}^{n} a_k (i\lambda)^k$ por $P(i\lambda)$ y sea $P_1(z)$ el resto correspondiente. Después, el quebrado $P_1(i\lambda)/P(i\lambda)$ se puede representar como una combinación lineal de "fracciones ordinarias" de la forma $1/(i\lambda - p)^k$ donde $p = \alpha + i\beta$, $\alpha < 0$, es la raíz del polinomio $P(z)$, y el grado k no supera la multiplicidad de esta raíz. Ya que

$$\frac{1}{(i\lambda - p)^k} = \int_0^{\infty} \frac{(-1)^k t^{k-1}}{(k-1)!} e^{-(i\lambda - p)t}\, dt,$$

se ve fácilmente, que cualquier combinación lineal de expresiones de la forma $1/(i\lambda - p)^k$ puede ser representada por la integral $\int_0^{\infty} e^{-i\lambda t} c(t)\, dt$ en la cual $c(t)$ es una combinación lineal de las funciones correspondientes de la forma

$$\frac{(-1)^k t^{k-1}}{(k-1)!} e^{pt} \quad 0 \leqslant t < \infty.$$

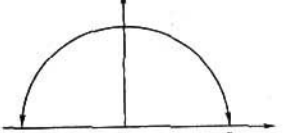

Fig. 25.

El lema queda demostrado.

Buscaremos una función $\varphi_0(i\lambda)$ en forma de la expresión (2.11) determinando los coeficientes a_k, $k = 0, \ldots, n-1$ de modo que la función $\psi(\lambda)$, ligada con $\varphi_0(i\lambda)$ por la igualdad (2.9) sea analítica en el semiplano superior de la variable compleja $\operatorname{Im}\lambda \geqslant 0$.

Debido a que la densidad espectral $f(\lambda) = \frac{|P(i\lambda)|^2}{|Q(i\lambda)|^2}$ tiene de polos en el semiplano superior las raíces $-iq_1, \ldots, -iq_n$ del polinomio $Q(i\lambda)$, entonces la función

$$\psi(\lambda) = \frac{e^{i\lambda\tau}P(i\lambda) - \sum_{k=0}^{n-1} a_k (i\lambda)^k}{Q(i\lambda)} \frac{P(-i\lambda)}{Q(-i\lambda)}$$

para cualesquiera coeficientes a_k, $k = 0, \ldots, n$, será analítica en el semiplano superior excluyendo puede ser los polos simples en los puntos $\lambda = -iq_1, \ldots, -iq_n$. Si determinamos los coeficientes a_k, $k = 0, \ldots, n-1$, del sistema de ecuaciones diferenciales

$$\sum_{k=0}^{n-1} a_k q_j^k = e^{q_j\tau}P(q_j), \qquad j = 1, \ldots, n, \tag{2.13}$$

que, para los diferentes $q_1, \ldots, q_n$, tiene un determinante distinto de cero

$$\begin{vmatrix} 1 & q_1 & q_1^2 & \ldots & q_1^{n-1} \\ 1 & q_2 & q_2^2 & \ldots & q_2^{n-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & q_n & q_n^2 & \ldots & q_n^{n-1} \end{vmatrix},$$

entonces la función correspondiente $\psi(\lambda)$ será analítica en el semiplano superior. Ya que $|\psi(\lambda)|$ decrece para $|\lambda| \to \infty$ como $|\lambda|^{-n+m-1}$, entonces según el lema 1 la transformación de Fourier de la función $\psi(\lambda)$ se convierte en cero en el semieje negativo $t \leqslant 0$.

Evidentemente, la función $\varphi_0(i\lambda)$ determinada por la igualdad (2.11) satisface la condición $\int_{-\infty}^{\infty} |\varphi_0(i\lambda)|^2 f(\lambda)\, d\lambda < \infty$. Según el lema 2, esta función está representada por la fórmula (2.12) de la cual se ve directamente, que $\varphi_0(i\lambda)$ pertenece al espacio L_0.

En efecto, la densidad espectral $f(\lambda)$ decrece para $|\lambda| \to \infty$ como $|\lambda|^{-2(n-m)}$, y las funciones determinadas sucesivamente por el paso límite

$$\varphi(\lambda) = \lim_{h \to -0} (i\lambda)^{k-1} \frac{e^{i\lambda(t+h)} - e^{i\lambda t}}{h} = (i\lambda)^k e^{i\lambda t}, \qquad k = 0, \ldots, n-m-1,$$

todas pertenecen a L_0, cuando $t \leqslant 0$ (para $t = 0$ tenemos $\varphi(\lambda) = (i\lambda)^k$, $k = 0, \ldots, n-m-1$); también pertenecen al espacio L_0 las funciones del tipo

$$\varphi(\lambda) = \int_0^\infty e^{-i\lambda t} c(t)\, dt = \int_{-\infty}^0 e^{i\lambda t} c(-t)\, dt,$$

donde la función $c(t)$ es continua e integrable absolutamente, en la semirecta $0 \leqslant t < \infty$ (véase el punto 2 del § 7).

Así, pues, hemos hallado que la función $\varphi_0(i\lambda) \in L_0$ satisface la condición (2.8) y, por consiguiente, la aproximación óptima para el valor $\xi(\tau)$ con las magnitudes $\eta \in L_0$ es

$$\hat{\xi} = \int_{-\infty}^{\infty} \varphi_0(i\lambda)\, d\Phi(\lambda) = \sum_{k=0}^{n-m-1} c_k \xi^{(k)}(0) + \int_{-\infty}^{0} c(-s)\, \xi(s)\, ds.$$

Se comprueba fácilmente que

$$\hat{\xi}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \varphi_0(i\lambda)\, d\Phi(\lambda) = \sum_{k=0}^{n-m-1} c_k \xi^{(k)}(t) + \int_{-\infty}^{t} c(t-s)\, \xi(s)\, ds \quad (2.14)$$

es la aproximación óptima del futuro valor $\xi(t+\tau)$ con las magnitudes $\eta \in H_{(-\infty,\, t)}$. En resumen hemos obtenido el siguiente resultado.

Teorema 1. *La aproximación óptima* $\xi(t+\tau)$ *con las magnitudes* $\eta \in H_{(-\infty,\, t)}$ *se da por la fórmula* (2.14) *en la cual la función* $\varphi_0(i\lambda)$ *está determinada por la igualdad* (2.11) *donde los coeficientes* $a_0, \ldots, a_{n-1}$ *satisfacen al sistema de ecuaciones lineales* (2.13).

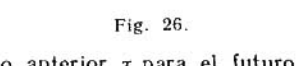

Fig. 26.

De tal modo el pronóstico lineal óptimo del proceso aleatorio estacionario $\xi(t)$, $-\infty < t < \infty$, en el tiempo anterior τ para el futuro puede realizarse por el sistema lineal con la función de transmisión

$$\varphi_0(p) = \sum_{k=0}^{n-1} a_k p^k / P(p).$$

2. Filtrado lineal (valuación del valor medio). Examinemos el proceso aleatorio $\xi = \xi(t)$ de la forma

$$\xi(t) = \theta(t) + \Delta(t), \quad (2.15)$$

donde $\theta = \theta(t)$ es una función de t (que en adelante se llama « señal ») y $\Delta = \Delta(t)$ es un proceso aleatorio con un valor medio nulo $\mathbf{M}\Delta(t) = 0$ (llamado en adelante «ruido aleatorio»). Examinemos la tarea sobre la separación de la señal $\theta = \theta(t)$ sobre el fondo del ruido aleatorio

$\Lambda = \Delta(t)$, más exactamente, la tarea sobre la valuación de la función $\theta = \theta(t)$ según la trayectoria correspondiente del proceso aleatorio $\xi(t) = \theta(t) + \Delta(t)$, $t \in T$, en un segmento T, considerando que las señales posibles $\theta = \theta(t)$ tienen la forma

$$\theta(t) = \sum_{k=1}^{m} \alpha_k \theta_k(t), \tag{2.16}$$

donde $\theta_1(t), \ldots, \theta_m(t)$ son ciertas funciones dadas linealmente independientes, y los coeficientes $\alpha_1, \ldots, \alpha_m$ pueden tomar cualesquiera valores reales.

Para la solución de esta tarea valuemos los coeficientes desconocidos $\alpha_1, \ldots, \alpha_m$ en la expresión (2.16).

Designemos por $H(T)$ el conjunto de todas las combinaciones lineales $\eta = \sum_h c_h \xi(t_h)$ y sus límites medios cuadráticos

$$(\eta = \text{l.i.m.} \sum c_h \xi(t_h), \text{ donde } t_h \in T.$$

Denominemos a las magnitudes $\eta_1, \ldots, \eta_m \in H(T)$ *valuaciones no desplazadas linealmente* de los coeficientes $\alpha_1, \ldots, \alpha_m$ si los valores medios $\mathbf{M}\eta_k$, $k = 1, \ldots, m$, para cualesquiera valores posibles $\alpha_1, \ldots, \alpha_m$ satisfacen la condición

$$\mathbf{M}\eta_k = \alpha_k, \quad k = 1, \ldots, m. \tag{2.17}$$

Ejemplo (valuación de los cuadrados mínimos).

Supongamos que la función desconocida $\theta = \theta(t)$ y la función aleatoria $\xi = \xi(t)$ considerada son integrables en el cuadrado sobre el segmento $T = [a, b]$. Para tales funciones cualesquiera $x_1 = x_1(t)$ y $x_2 = x_2(t)$ pongamos

$$(x_1, x_2) = \int_a^b x_1(t)\, x_2(t)\, dt.$$

Según la trayectoria existente $\xi(t)$, $a \leqslant t \leqslant b$, determinemos las magnitudes $\eta_1, \ldots, \eta_m$ de la condición

$$\int_a^b \left[\xi(t) - \sum_{k=1}^{n} \eta_k \theta_k(t)\right]^2 dt = \min_{y_1, \ldots, y_m} \int_a^b \left[\xi(t) - \sum_{k=1}^{n} y_k \theta_k(t)\right]^2 dt,$$

donde el número mínimo se toma por todos los valores reales posibles $y_1, \ldots, y_m$. Se trata del mínimo de la forma cuadrática no negativa

$$R(y_1, \ldots, y_m) = (\xi, \xi) - 2\sum_{k=1}^{m} y_k (\theta_k \xi) + \sum_{k,j=1}^{m} y_k y_j (\theta_k \theta_j)$$

de las variables $(y_1, \ldots, y_m)$. Este mínimo se alcanza en el punto $(\eta_1, \ldots, \eta_m)$ en el cual

$$\frac{\partial}{\partial y_k} R(\eta_1, \ldots, \eta_m) = -2\left[(\theta_k, \xi) + \sum_{j=1}^{m} \eta_j (\theta_k, \theta_j)\right] = 0, \quad k = 1, \ldots, m,$$

lo que nos da el sistema de ecuaciones lineales con relación a $\eta_1, \ldots, \eta_m$

$$\sum_{j=1}^{m} \eta_j (\theta_k, \theta_j) = (\xi, \theta_k), \quad k = 1, \ldots, m. \tag{2.18}$$

Para las funciones linealmente independientes $\theta_1(t), \ldots, \theta_m(t)$, la matriz $D = \{(\theta_k, \theta_j)\}$ de los coeficientes

$$(\theta_k \theta_j) = \int_a^b \theta_k(t)\, \theta_j(t)\, dt, \quad k, j = 1, \ldots, m,$$

del sistema de ecuaciones (2.18) es no degenerado y la solución de este sistema es

$$\eta_k = \sum_{j=1}^{m} \sigma_{kj} (\theta_j, \xi), \quad k = 1, \ldots, m, \tag{2.19}$$

donde σ_{kj} son los elementos de la matriz $\{\sigma_{kj}\}$ inversa nos da la matriz D.

Para cualquier función $\theta_j(t)$ la integral $(\theta_j, \xi) = \int_a^b \theta_j(t)\, \xi(t)\, dt$ es el límite medio cuadrático de las sumas integrales correspondientes, que representan en sí las combinaciones lineales de la forma $\sum_k c_k \xi(t_k)$. Por consiguiente, todas las magnitudes (θ_j, ξ) en la expresión (2.19) y junto con ellas, también las magnitudes $\eta_1, \ldots, \eta_m$ entran en el espacio lineal $H(T)$.

Además de esto, ya que el valor medio $\mathbf{M}\xi(t)$ es $\theta(t) = \sum_{k=1}^{m} \alpha_k \theta_k(t)$, tenemos que

$$\mathbf{M}(\theta_j, \xi) = \mathbf{M} \int_a^b \theta_j(t)\, \xi(t)\, dt = \int_a^b \theta_j(t)\, [\mathbf{M}\xi(t)]\, dt =$$

$$= \int_a^b \theta_j(t)\, \theta(t)\, dt = (\theta_j, \theta),$$

y se ve que los valores medios

$$\mathbf{M}\eta_k = \sum_{k=1}^{m} \sigma_{kj} (\theta_j, \theta), \quad k = 1, \ldots, m,$$

se obtienen de la fórmula (2.19) sustituyendo la función $\xi = \xi(t)$ por la función $\theta = \theta(t)$. Por consiguiente, estos valores dan el punto del mínimo de la expresión

$$\int_a^b \left| \theta(t) - \sum_{k=1}^{m} y_k \theta_k(t) \right|^2 dt \text{ y, ya que para } \theta(t) = \sum_{k=1}^{m} \alpha_k \theta_k(t)$$

el mínimo igual a cero, se alcanza para $y_k = \alpha_k$, $k = 1, \ldots, m$, de aquí deducimos que

$$\mathbf{M}\eta_k = \alpha_k, \quad k = 1, \ldots, m.$$

De tal modo, las magnitudes $\eta_1, \ldots, \eta_m$ determinadas por la fórmula (2.19) son las valuaciones lineales no desplazadas de los coeficientes desconocidos $\alpha_1, \ldots, \alpha_m$ (corrientemente se las llama *valuaciones de los cuadrados mínimos*).

Llamemos valuaciones lineales no desplazadas $\hat{\alpha}_1, \ldots, \hat{\alpha}_m$ de los coeficientes $\alpha_1, \ldots, \alpha_m$ a las *valuaciones óptimas*, si

$$\| \alpha_k - \hat{\alpha}_k \| = \min_{\eta \in L_k} \| \alpha_k - \eta \|, \quad k = 1, \ldots, m, \qquad (2.20)$$

donde el mínimo se toma por todas las valuaciones lineales no desplazadas η para el coeficiente correspondiente α_k (cuyo conjunto lo designamos por L_k).

Por definición L_k es el hiperplano (en el subespacio lineal $H(T)$) de todas las magnitudes $\eta \in H(T)$ que satisfacen la condición $\mathbf{M}\eta = \alpha_k$. Por eso (véase el lema sobre la perpendicular, pág. 253), las valuaciones óptimas no desplazadas $\hat{\alpha}_k \in L_k$ se determinan unívocamente por la condición

$$\langle \alpha_k - \hat{\alpha}_k, \ \eta - \hat{\alpha}_k \rangle = 0 \text{ para todos los } \eta \in L_k, \quad k = 1, \ldots, m. \qquad (2.21)$$

Intentemos describir más explícitamente el espacio $H(T)$ y los hiperplanos L_k, $k = 1, \ldots, m$ suponiendo que en la expresión (2.15) $\Delta = \Delta(t)$ es un proceso aleatorio estacionario con la densidad espectral $f(\lambda)$ que admite la representación espectral

$$\Delta(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \, d\Phi_0(\lambda), \qquad (2.22)$$

y la función $\theta = \theta(t)$ es la transformación de Fourier de $\tilde{\theta} = \tilde{\theta}(\lambda)$:

$$\theta(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \tilde{\theta}(\lambda) \, d\lambda, \qquad (2.23)$$

donde la función $\widetilde{\theta}(\lambda)$ satisface la condición

$$\int_{-\infty}^{\infty} \frac{|\widetilde{\theta}(\lambda)|^2}{f(\lambda)}\, d\lambda < \infty. \tag{2.24}$$

Las fórmulas (2.22), (2.23) nos permiten representar el proceso aleatorio examinado $\xi(t) = \theta(t) + \Delta(t)$ por la integral estocástica

$$\xi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\, d\Phi(\lambda) \tag{2.25}$$

(compárese con la representación espectral (7.4), cap. III), donde

$$\Phi(\lambda) = \int_{-\infty}^{\lambda} \widetilde{\theta}(\lambda)\, d\lambda + \Phi_0(\lambda)$$

es un proceso aleatorio con los incrementos no correlacionados, que tienen el valor medio

$$\mathbf{M}\Phi(\lambda) = \int_{-\infty}^{\lambda} \widetilde{\theta}(\lambda)\, d\lambda$$

y la densidad de estructura $f(\lambda)$.

Igualmente que para los procesos estacionarios, examinaremos el espacio lineal $H(T)$ de todas las magnitudes η, representadas por la integral estocástica

$$\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\, d\Phi(\lambda), \tag{2.26}$$

donde las funciones $\varphi(\lambda)$ entran en el espacio correspondiente $L_T(f)$, con ello existe la integral $\int_{-\infty}^{\infty} \varphi(\lambda)\, \widetilde{\theta}(\lambda)\, d\lambda$ (véase el punto 2, del § 6).

Hablando en general, la existencia de la integral $\int_{-\infty}^{\infty} \varphi(\lambda)\, \widetilde{\theta}(\lambda)\, d\lambda$ es la condición complementaria para las funciones $\varphi(\lambda) \in L_T(f)$, ahora bien, para las suposiciones hechas anteriormente, ella se deduce automáticamente de la condición general $\int_{-\infty}^{\infty} |\varphi(\lambda)|^2 f(\lambda)\, d\lambda <$

$< \infty$ ya que

$$\int_{-\infty}^{\infty} |\varphi(\lambda)\overline{\theta}(\lambda)|\,d\lambda = \int_{-\infty}^{\infty} \left|\varphi(\lambda)\sqrt{f(\lambda)}\,\frac{\overline{\theta}(\lambda)}{\sqrt{f(\lambda)}}\right| d\lambda \leqslant$$

$$\leqslant \sqrt{\int_{-\infty}^{\infty} |\varphi(\lambda)|^2 f(\lambda)\,d\lambda}\,\sqrt{\int_{-\infty}^{\infty} \frac{|\overline{\theta}(\lambda)|^2}{f(\lambda)}\,d\lambda} < \infty.$$

De tal modo, *el espacio lineal H (T) está compuesto por todas las magnitudes*

$$\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\,d\Phi(\lambda), \quad \varphi(\lambda) \in L_T(f).$$

Recordemos que para las integrales estocásticas (2.26)

$$\mathbf{M}\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\,\overline{\theta}(\lambda)\,d\lambda, \tag{2.27}$$

y

$$\langle \eta_1 - \mathbf{M}\eta_1,\ \eta_2 - \mathbf{M}\eta_2 \rangle = \langle \varphi_1,\ \varphi_2 \rangle, \tag{2.28}$$

donde φ_1, φ_2 son las funciones correspondientes del espacio $L_T(f)$ y

$$\langle \varphi_1,\ \varphi_2 \rangle = \int_{-\infty}^{\infty} \varphi_1(\lambda)\,\overline{\varphi_2(\lambda)}\,f(\lambda)\,d\lambda.$$

Recurramos a las funciones

$$\psi_1(\lambda),\ \ldots,\ \psi_m(\lambda) \in L_T(f),$$

que satisfacen las ecuaciones integrales [1])

$$\int_{-\infty}^{\infty} e^{-it\lambda}\psi_k(\lambda)\,f(\lambda)\,d\lambda = \theta_k(t), \quad t \in T, \quad k = 1,\ \ldots,\ m. \tag{2.29}$$

Como se deduce de la fórmula (2.27) que determina el hiperplano L_k la condición $\mathbf{M}\eta = \alpha_k$ para las magnitudes $\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\,d\Phi(\lambda)$

[1]) Señalemos que tales funciones $\psi_1, \ldots, \psi_m$ existen siempre, teniendo para cada $k = 1, \ldots, m$ la única solución $\psi_k(\lambda)$ de la ecuación (2.29), que pertenece al espacio $L_T(f)$ (con relación a estas ecuaciones véase, por ejemplo, el libro de Yu. A. Rozanov, «Procesos aleatorios estacionarios», Fismatguiz, Moscú, 1963, pág. 190).

significa que

$$\int_{-\infty}^{\infty} \varphi(\lambda)\,\overline{\theta}(\lambda)\,d\lambda = \int_{-\infty}^{\infty} \varphi(\lambda)\left[\sum_{j=1}^{m} \alpha_j \overline{\psi_j(\lambda)}\right] f(\lambda)\,d\lambda = \sum_{j=1}^{m} \alpha_j \langle \varphi, \psi_j\rangle \equiv \alpha_k \tag{2.30}$$

para todos los $\alpha_1, \ldots, \alpha_m$, lo que, en vista de la arbitrariedad de estos coeficientes, es equivalente a la condición

$$\langle \varphi, \psi_j\rangle = \begin{cases} 1 \text{ para } j = k, \\ 0 \text{ para } j \neq k. \end{cases} \tag{2.31}$$

Hemos hallado que *cada uno de los hiperplanos* L_k *(de las valuaciones lineales no desplazadas* $\eta = \int_{-\infty}^{\infty} \varphi(\lambda)\,d\Phi(\lambda)$ *para los coeficientes incógnitos* α_k*) se determina por el sistema de ecuaciones (2.31) en el cual las funciones* $\psi_1(\lambda), \ldots, \psi_m(\lambda)$ *satisfacen las ecuaciones integrales* (2.29).

Volvamos a la condición (2.21) que determina la valuación óptima $\hat{\alpha}_k = \int_{-\infty}^{\infty} \varphi_{0k}(\lambda)\,d\Phi(\lambda)$ para cada $k = 1, \ldots, m$. De la fórmula general (2.28) se desprende que esta condición para la función correspondiente $\varphi_{0k}(\lambda) \in L_k$ significa lo siguiente:

$$\langle \varphi_{0k},\ \varphi - \varphi_{0k}\rangle = 0 \tag{2.32}$$

para todos los $\varphi(\lambda) \in L_T(f)$ que satisfacen las ecuaciones (2.31).

Introduzcamos la matriz $D = \{d_{kj}\}$ con los elementos

$$d_{kj} = \langle \psi_k,\ \psi_j\rangle = \int_{-\infty}^{\infty} \psi(\lambda)\,\overline{\psi_j(\lambda)}\,f(\lambda)\,d\lambda. \tag{2.33}$$

Ya que para las funciones linealmente independientes $\theta_1(t), \ldots, \theta_m(t)$ las funciones $\psi_1(\lambda), \ldots, \psi_m(\lambda)$ determinades por las ecuaciones (2.29), también son linealmente independientes, la matriz D no está degenerada y existe la matriz inversa $D^{-1} = \{\sigma_{kj}\}$:

$$\sum_{j=1}^{m} \sigma_{ij} d_{jh} = \begin{cases} 1 \text{ para } k = i, \\ 0 \text{ para } k \neq i. \end{cases}$$

Mostremos que las funciones desconocidas $\varphi_{0k}(\lambda)$ son

$$\varphi_{0k}(\lambda) = \sum_{j=1}^{m} \sigma_{kj}\psi_j(\lambda), \qquad k = 1, \ldots, m. \tag{2.34}$$

En efecto, ya que todas las funciones $\psi_j(\lambda)$, $j = 1, \ldots, m$, pertenecen al espacio lineal $L_T(f)$, también pertenecen a él las funciones

$\varphi_{0k}(\lambda)$, $k = 1, \ldots, m$. Cada una de éstas satisface la condición respectivo (2.31) a saber:

$$\langle \varphi_{0k}, \psi_j \rangle = \sum_{i=1}^{m} \sigma_{ki} \langle \psi_i, \psi_j \rangle = \sum_{i=1}^{m} \sigma_{ki} d_{ij} = \begin{cases} 1 \text{ para } j = k, \\ 0 \text{ para } j \neq k. \end{cases}$$

Además de esto, para cada $k = 1, \ldots, m$ tiene lugar la igualdad (2.32), ya que para cualquier función $\varphi(\lambda)$ que satisface la condición (2.31).

$$\langle \varphi_{0k}, \varphi \rangle = \sum_{j=1}^{m} \sigma_{kj} \langle \psi_j, \varphi \rangle = \sigma_{kk}$$

y al mismo tiempo

$$\langle \varphi_{0k}, \varphi_{0k} \rangle = \sum_{i,\, j=1}^{m} \sigma_{ki} d_{ij} \sigma_{jk} = \sigma_{kk}.$$

Señalemos que la matriz $D^{-1} = \{\sigma_{kj}\}$ es la matriz correlativa de los errores $\hat{\alpha}_k - \alpha_k$, $k = 1, \ldots, m$ en las valuaciones óptimas de los coeficientes $\alpha_1, \ldots, \alpha_m$, a saber:

$$\sigma_{kj} = \langle \varphi_{0k}, \varphi_{0j} \rangle = \mathbf{M}(\hat{\alpha}_k - \alpha_k)(\hat{\alpha}_j - \alpha_j), \quad k, j = 1, \ldots, m. \tag{2.35}$$

Enunciemos el resultado que hemos obtenido.

Teorema 2. *Las valuaciones óptimas lineales* $\hat{\alpha}_1, \ldots, \hat{\alpha}_m$ *de los coeficientes desconocidos* $\alpha_1, \ldots, \alpha_m$ *se dan por la expresión*

$$\alpha_k = \int_{-\infty}^{\infty} \varphi_{0k}(\lambda)\, d\Phi(\lambda), \quad k = 1, \ldots, m,$$

donde las funciones $\varphi_{0k}(\lambda)$, $k = 1, \ldots, m$, *están ligadas por las fórmulas (2.34) con la solución* $\psi_k(\lambda) \in L_T(f)$, $k = 1, \ldots, m$ *de las ecuaciones integrales (2.29).*

Señalemos que algunas veces se puede valuar s i n e r r o r [1]) los coeficientes desconocidos α_k, $k = 1, \ldots, m$.

Ejemplo (ley de los grandes números). Sea

$$\theta(t) = \alpha, \quad 0 \leqslant t < \infty,$$

donde α es una constante desconocida. Entonces

$$\alpha = \text{l.i.m.}_{n \to \infty} \frac{1}{n} \int_0^n \xi(t)\, dt.$$

[1]) Así ocurrirá siempre al infringir la condición (2.24). Véase, por ejemplo, el libro de I. A. I b r a g u i m o v, Y u. A. R o z a n o v «Procesos aleatorios de Gauss», Naúka, Moscú, 1970.

En efecto, $\frac{1}{n}\int_0^n \xi(t)\,dt-\alpha=\frac{1}{n}\int_0^n \Delta(t)\,dt$, donde para el proceso estacionario $\Delta(t)=\int_{-\infty}^{\infty} e^{i\lambda t}\,d\Phi_0(\lambda)$ tenemos

$$\frac{1}{n}\int_0^n \Delta(t)\,dt=\int_{-\infty}^{\infty}\left[\frac{1}{n}\int_0^n e^{i\lambda t}\,dt\right]d\Phi_0(\lambda)=\int_{-\infty}^{\infty}\frac{e^{i\lambda n}-1}{i\lambda n}\,d\Phi_0(\lambda),$$

y se ve fácilmente, que para $n\to\infty$

$$\left\|\frac{1}{n}\int_0^n \xi(t)\,dt-\alpha\right\|^2=\int_{-\infty}^{\infty}\left|\frac{e^{i\lambda n}-1}{i\lambda n}\right|^2 f(\lambda)\,d\lambda=\int_{-\infty}^{\infty}\frac{4\,\mathrm{sen}^2\,\lambda n}{\lambda^2 n^2}\,f(\lambda)\,d\lambda\to 0.$$

§ 3. ESPERANZAS MATEMATICAS CONDICIONALES Y ALGUNAS TAREAS DE PRONOSTICACIÓN Y DE FILTRADO

1. Una vez más sobre las esperanzas matemáticas condicionales. Examinemos una familia arbitraria de magnitudes aleatorias $\eta(t)$, $t\in T$, dependientes del parámetro t y la magnitud aleatoria ξ, que de algún modo está ligada con las magnitudes $\eta(t)$, $t\in T$.

Para mayor claridad nos podemos imaginar una prueba compleja en la que al principio se observan las magnitudes $\eta(t)$ $t\in T$, y luego, en dependencia de los valores $\eta(t)$, $t\in T$, se realiza una prueba complementaria, en la cual se observa la magnitud aleatoria ξ; en otras palabras, «el mecanismo de aleatoriedad» actúa sucesivamente: al principio, las magnitudes $\eta(t)$, $t\in T$, toman tales o cuales valores con cierta distribución de probabilidades y, luego, la magnitud ξ toma tal o cual valor con cierta distribución (condicional) de probabilidades, en dependencia de aquellas magnitudes.

En estas condiciones la esperanza matemática de la magnitud ξ para los valores fijados de $\eta(t)$, $t\in T$ (designémosla por $\mathbf{M}\{|\xi|\eta\}$) depende de las magnitudes aleatorias $\eta(t)$, $t\in T$, y en este sentido es una magnitud aleatoria. Parece intuitivamente que se puede determinar el valor medio $\mathbf{M}\xi$ tomando al principio $M\{\xi|\eta\}$ para los valores fijados $\eta(t)$, $t\in T$, y luego tomando la esperanza matemática $\mathbf{M}[M\{\xi|\eta\}]$:

$$\mathbf{M}\xi=\mathbf{M}[\mathbf{M}\{\xi|\eta\}]. \tag{3.1}$$

Además de esto, para cualquier magnitud aleatoria $\varphi(\eta)$ determinada unívocamente por los valores $\eta(t)$, $t \in T$, se deberá cumplir la igualdad

$$\mathbf{M}\{\varphi(\eta)\cdot\xi \mid \eta\} = \varphi(\eta)\cdot\mathbf{M}\{\xi \mid \eta\}, \tag{3.2}$$

ya que para los valores fijados $\eta(t)$, $t \in T$, la magnitud $\varphi(\eta)$ es constante.

Designemos por L el conjunto de tales magnitudes $\varphi(\eta)$ para las cuales $\mathbf{M}\varphi^2(\eta) < \infty$. Supongamos que para cada uno de los valores fijados $\eta(t)$, $t \in T$, se tiene la distribución condicional de probabilidades correspondiente, con ello la esperanza matemática condicional $\mathbf{M}\{\xi \mid \eta\}$ posee las propiedades (3.1) y (3.2).

Examinemos la magnitud ξ, $\mathbf{M}\xi^2 < \infty$. La magnitud correspondiente $\hat{\xi} = \mathbf{M}(\xi \mid \eta)$ entra en L, ya que según la desigualdad general para los momentos

$$\hat{\xi}^2 = [\mathbf{M}(\xi \mid \eta)]^2 \leqslant \mathbf{M}(\xi^2 \mid \eta),$$

donde de acuerdo con la fórmula (3.1) aplicada a la magnitud ξ^2,

$$\mathbf{M}[\mathbf{M}(\xi^2 \mid \eta)] = \mathbf{M}\xi^2$$

y, por consiguiente, $\mathbf{M}\hat{\xi}^2 \leqslant \mathbf{M}\xi^2 < \infty$. Según la fórmula (3.2) $\mathbf{M}\{\varphi(\eta)\cdot\xi \mid \eta\} = \varphi(\eta)\cdot\hat{\xi}$, de modo que la igualdad (3.1) para $\varphi(\eta)\,\xi$ da

$$\mathbf{M}[\mathbf{M}\{\varphi(\eta)\cdot\xi \mid \eta\}] = \mathbf{M}[\varphi(\eta)\cdot\xi] = \langle\varphi(\eta), \xi\rangle$$

y al mismo tiempo

$$\mathbf{M}[\mathbf{M}\{\varphi(\eta)\cdot\xi \mid \eta\}] = \mathbf{M}[\varphi(\eta)\cdot\hat{\xi}] = \langle\varphi(\eta), \hat{\xi}\rangle,$$

de donde obtenemos que

$$\langle\xi - \hat{\xi}, \varphi(\eta)\rangle = 0 \text{ para todos los valores } \varphi(\eta) \in L. \tag{3.3}$$

Pero antes, en el § 1 fue mostrado que existía, una sola magnitud $\hat{\xi} \in L$ que satisfacía la condición (3.3): $\hat{\xi} = \mathbf{M}\{\xi \mid \eta\}$ *esto significa geométricamente la proyección de la magnitud* ξ *sobre el espacio lineal* L *de todas las magnitudes* $\varphi(\eta)$, $\mathbf{M}\varphi^2(\eta) < \infty$ *que son funciones de los valores* $\eta(t)$, $t \in T$.

Partiendo de las esperanzas matemáticas condicionales, se puede determinar la distribución condicional de probabilidades ya que

$$P\{x' \leqslant \xi \leqslant x'' \mid \eta\} = \mathbf{M}\{\varphi(\xi) \mid \eta\}, \tag{3.4}$$

donde la función $\varphi(x)$ tiene la forma

$$\varphi(x) = \begin{cases} 1 \text{ para } x' \leqslant x \leqslant x'' \\ 0 \text{ para los restantes } x. \end{cases}$$

Ejemplo. Sean ξ y η magnitudes aleatorias discretas, que tienen una distribución conjunta de probabilidades $P_{\xi\eta}(x, y)$, $-\infty < x, y < \infty$. Hallemos la distribución condicional de probabilidades $P_\xi(x \mid \eta)$ de la magnitud ξ para el valor fijado de η.

Para la magnitud aleatoria $\varphi(\xi)$ que es función de ξ, la esperanza matemática con relación a la distribución de probabilidades $P_\xi(x \mid \eta)$ para el valor fijado de η es

$$\mathbf{M}\{\varphi(\xi) \mid \eta\} = \sum_{-\infty}^{\infty} \varphi(x) P_\xi(x \mid \eta).$$

Para la función limitada $\psi = \psi(y)$ tendremos

$$\mathbf{M}\{\varphi(\xi)\psi(\eta) \mid \eta\} = \sum_{-\infty}^{\infty} \varphi(x)\psi(\eta) P_\xi(x \mid \eta).$$

Tomando las esperanzas matemáticas de la parte izquierda y derecha, por una parte obtenemos

$$\mathbf{M}[\mathbf{M}\{\varphi(\xi)\psi(\eta) \mid \eta\}] = \mathbf{M}\varphi(\xi)\psi(\eta) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \varphi(x)\psi(y) P_{\xi\eta}(x, y),$$

y de otra parte,

$$\mathbf{M}[\mathbf{M}\{\varphi(\xi)\psi(\eta) \mid \eta\}] = \mathbf{M}\left[\sum_{-\infty}^{\infty} \varphi(x)\psi(\eta) P_\xi(x \mid \eta)\right] =$$

$$= \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \varphi(x)\psi(y) P_\xi(x \mid y) P_\eta(y)$$

(donde $P_\eta(y)$ es la distribución de la magnitud η), lo que da la siguiente igualdad:

$$\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \varphi(x)\psi(y) P_\xi(x \mid y) P_\eta(y) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty} \varphi(x)\psi(y) P_{\xi\eta}(x, y).$$

De aquí, suponiendo que

$$\varphi(x) = \begin{cases} 1 \text{ para } x = x_0, \\ 0 \text{ para } x \neq x_0; \end{cases}$$

$$\psi(y) = \begin{cases} 1 \text{ para } y = y_0, \\ 0 \text{ para } y \neq y_0 \end{cases}$$

para cualesquiera x_0, y_0, obtenemos, que

$$P_\xi(x_0 \mid y_0) P_\eta(y_0) = P_{\xi\eta}(x_0, y_0).$$

De tal modo, para cualesquiera valores de y, que toma la magnitud η con probabilidad positiva.

$$P_\xi(x \mid y) = \frac{P_{\xi\eta}(x, y)}{P_\eta(y)} \tag{3.5}$$

(compárese con (3.16), cap. I).

Ejemplo. Sea η una magnitud aleatoria discreta con distribución de probabilidades $P_\eta(y)$. Supongamos que la magnitud ξ para cada valor fijado η tiene la densidad condicional de probabilidades $p_\xi(x \mid \eta)$. Para

$$\varphi(x) = \begin{cases} 1 \text{ para } x' \leqslant x \leqslant x'', \\ 0 \text{ para } x < x',\ x > x''; \end{cases}$$

$$\psi(y) = \begin{cases} 1 \text{ para } y = y_0, \\ 0 \text{ para } y \neq y_0 \end{cases}$$

tenemos

$$\mathbf{M}[\mathbf{M}\{\psi(\eta)\varphi(\xi) \mid \eta\}] = \mathbf{M}\left[\psi(\eta)\int_{x'}^{x''} p_\xi(x \mid \eta)\,dx\right] =$$

$$= \sum_y \left[\psi(y)\int_{x'}^{x''} p_\xi(x \mid y)\,dx\right] P_\eta(y) = \int_{x'}^{x''} p_\xi(x \mid y_0)\,dx \cdot P_\eta(y_0).$$

Ahora bien, para cualquier y_0 la fórmula (3.1) aplicada a la magnitud $\psi(\eta)\varphi(\xi)$ nos da a la vez:

$$\mathbf{M}[\mathbf{M}\{\psi(\eta)\varphi(\xi) \mid \eta\}] = \mathbf{M}\psi(\eta)\varphi(\xi) = \\ = \mathbf{P}\{x' \leqslant \xi \leqslant x'',\ \eta = y_0\},$$

de modo que en total obtenemos

$$P\{x' \leqslant \xi \leqslant x'',\ \eta = y\} = \int_{x'}^{x''} p_\xi(x \mid y)\,dx P_\eta(y) \qquad (3.6)$$

para todas los x', x'' e y.

Ejemplo. Sea ξ una magnitud discreta con distribución de probabilidades $P_\xi(x)$, $-\infty < x < \infty$ y la magnitud aleatoria η para cada valor fijado de ξ tiene la densidad de probabilidad $p_\eta(y \mid \xi)$, $-\infty < y < \infty$. Hallemos la distribución conjunta de probabilidades de estas magnitudes y la distribución condicional $P_\xi(x \mid \eta)$ de la magnitud ξ respecto a η.

La probabilidad del suceso $\{y' \leqslant \eta \leqslant y''\}$ para la condición $\xi = x$ es

$$P\{y' \leqslant \eta \leqslant y'' \mid \xi = x\} = \int_{y'}^{y''} p_\eta(y \mid x)\,dy.$$

Según la fórmula (3.6)

$$P\{\xi = x,\ y' \leqslant \eta \leqslant y''\} = P_\xi(x) \int_{y'}^{y''} p_\eta(y \mid x)\,dy,$$

y por, consiguiente, la distribución conjunta se da como

$$\mathbf{P}\{x' \leqslant \xi \leqslant x'',\ y' \leqslant \eta \leqslant y''\} = \int_{y'}^{y''} \left[\sum_{x'}^{x''} P_\xi(x)\, p_\eta(y \mid x) \right] dy.$$

De aquí se deduce, en particular, que

$$\mathbf{P}\{y' \leqslant \eta \leqslant y''\} = \int_{y'}^{y''} \left[\sum_{-\infty}^{\infty} P_\xi(x)\, p_\eta(y \mid x) \right] dy,$$

y se ve que la magnitud aleatoria η tiene la densidad de probabilidad

$$p_\eta(y) = \sum_{-\infty}^{\infty} p_\eta(y \mid x)\, P_\xi(x).$$

Mostremos que la distribución condicional $P_\xi(x \mid y)$ de la magnitud aleatoria ξ para el valor fijo de $\eta = y$ se da por la llamada *fórmula de Bayes*:

$$P_\xi(x \mid y) = P_\xi(x)\, \frac{p_\eta(y \mid x)}{p_\eta(y)}. \tag{3.7}$$

En realidad, la fórmula (3.7) da la distribución de probabilidades, con relación a la cual el valor medio

$$\mathbf{M}\{\varphi(\xi)\, \eta\} = \sum_{-\infty}^{\infty} \varphi(x)\, P_\xi(x \mid \eta)$$

de las magnitudes de la forma $\varphi(\xi)$ satisface la condición (3.1):

$$\mathbf{M}[\mathbf{M}\{\varphi(\xi) \mid \eta\}] = \int_{-\infty}^{\infty} \mathbf{M}\{\varphi(\xi) \mid y\}\, p_\eta(y)\, dy =$$

$$= \sum_{-\infty}^{\infty} \varphi(x)\, P_\xi(x) \left[\int_{-\infty}^{\infty} p_\eta(y \mid x)\, dy \right] = \sum_{-\infty}^{\infty} \varphi(x)\, P_\xi(x) = \mathbf{M}\varphi(\xi),$$

y también la condición (3.2) y, como ya sabemos, esta condición determina unívocamente la magnitud $\mathbf{M}\{\varphi(\xi) \mid \eta\}$; en su conjunto (para distintos $\varphi(\xi)$ los valores medios $\mathbf{M}\{\varphi(\xi) \mid \eta\}$ determinan unívocamente la distribución de probabilidades correspondiente $P_\xi(x \mid \eta)$ (digamos, $\mathbf{M}\{\varphi(\xi) \mid \eta\} = \mathbf{P}_\xi(x_0 \mid \eta)$ para la función $\varphi(x)$ que es igual a 1 para $x = x_0$ e igual a cero para los restantes x).

Ejemplo. Sean ξ y η magnitudes aleatorias que tienen la densidad conjunta de probabilidad $p_{\xi\eta}(x, y)$.

Mostremos que la densidad condicional $p_\xi(x \mid \eta)$ de la magnitud ξ con relación a η tiene la forma

$$p_\xi(x \mid y) = \frac{p_{\xi\eta}(x, y)}{p_\eta(y)}, \quad -\infty < x < \infty, \tag{3.8}$$

donde

$$p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi\eta}(x, y)\, dx, \quad -\infty < y < \infty,$$

es la densidad de probabilidad de la magnitud η (compárese con (3.17), cap. I).

En efecto, la esperanza matemática de la magnitud del tipo $\varphi(\xi)$ con respecto a la distribución con la densidad de probabilidades indicada $p_\xi(x \mid \eta)$:

$$\mathbf{M}\{\varphi(\xi) \mid \eta\} = \int_{-\infty}^{\infty} \varphi(x)\, p_\xi(x \mid \eta)\, dx,$$

satisface la condición (3.1) a saber:

$$\mathbf{M}[\mathbf{M}\{\varphi(\xi) \mid \eta\}] = \int_{-\infty}^{\infty} [\varphi(x)\, p_\xi(x \mid y)\, dx]\, p_\eta(y)\, dy =$$

$$= \int_{-\infty}^{\infty} \varphi(x) \Big[\int_{-\infty}^{\infty} p_{\xi\eta}(x, y)\, dy \Big] dx = \int_{-\infty}^{\infty} \varphi(x)\, p_\xi(x)\, dx = \mathbf{M}\varphi(\xi),$$

de tal modo que $\mathbf{M}\{\varphi(\xi) \mid \eta\}$ da realmente la esperanza matemática condicional de las magnitudes $\varphi(\xi)$ con relación a η. Pero en su conjunto (para distintos $\varphi(\xi)$ los valores medios $\mathbf{M}\{\varphi(\xi) \mid \eta\}$ determinan unívocamente la distribución de probabilidades, en particular, según la fórmula (3.4).

$$\mathbf{P}\{x' \leqslant \xi \leqslant x'' \mid \eta\} = \int_{x'}^{x''} p_\xi(x \mid \eta)\, dx,$$

y de tal modo, la función $p_\xi(x \mid \eta)$, $-\infty < x < \infty$ dada por la igualdad (3.8) es la densidad condicional de probabilidad de la magnitud ξ para el valor fijo η.

Señalemos que la densidad condicional $p_\eta(y \mid \xi)$ se determina análogamente a (3.8), con ello tiene lugar la siguiente igualdad (compárese con la fórmula de Bayes (3.7)

$$p_\xi(x \mid y) = \frac{p_\eta(y \mid x)}{p_\eta(y)}\, p_\xi(x). \tag{3.9}$$

2. Papel de las probabilidades a posteriori en algunas tareas de pronosticación y filtrado.

Como hemos visto, las propiedades fundamentales de las esperanzas matemáticas condicionales expresadas por las relaciones (3.1) y (3.2) permiten dar la siguiente interpretación geométrica del valor medio condicional $\hat{\xi} = \mathbf{M}(\xi \mid \eta)$ de la magnitud ξ, $\mathbf{M}\xi^2 < \infty$, en relación a las

magnitudes $\eta(t)$, $t \in T$: $\hat{\xi} = \mathbf{M}(\xi \mid \eta)$ *es la proyección de* ξ *en el espacio* L *de todas las magnitudes de la forma* $\varphi(\eta)$ *que son funciones de los valores* $\eta(t)$, $t \in T$. Con esto *la magnitud* ξ *da la aproximación óptima de* ξ con las magnitudes $\varphi(\eta) \in L$:

$$\| \xi - \hat{\xi} \| = \min_{\varphi(\eta)} \| \xi - \varphi(\eta) \|.$$

Examinemos la tarea sobre la valuación del valor de ξ por las magnitudes de $\eta(t)$, $t \in T$, en el caso cuando sólo se tiene un número finito de valores posibles ξ. Para mayor claridad se puede suponer que ξ describe el estado del sistema que nos interesa, sobre el que sólo podemos juzgar por la observación de un proceso aleatorio $\eta = \eta(t)$.

En ejemplos sencillos nos podemos convencer fácilmente de que, hablando en general, la esperanza matemática $\mathbf{M}(\xi \mid \eta)$ no toma obligatoriamente uno de los valores posibles de ξ (digamos, $\xi = 0, 1, \ldots, N$) de modo que la magnitud $\mathbf{M}(\xi \mid \eta)$ no nos da directamente una indicación respecto a los estados posibles $0, 1, \ldots, N$ en que se encuentra precisamente el sistema.

En adelante, llamaremos a la magnitud $\varphi_0(\eta) \in L$ *la valuación óptima* del estado desconocido ξ, si la probabilidad de los errores, es decir, $\mathbf{P}\{\varphi_0(\eta) \neq \xi\}$ es mínima:

$$\mathbf{P}\{\varphi_0(\eta) \neq \xi\} = \min_{\varphi(\eta)} \mathbf{P}\{\varphi(\eta) \neq \xi\}. \tag{3.10}$$

Está claro que para hallar la valuación óptima es suficiente limitarse a las magnitudes $\varphi(\eta) \in L$, cuyos valores posibles son los mismos que para las magnitudes ξ (es decir, según la suposición hecha son $0, 1, \ldots, N$).

Suponiendo como anteriormente, que existe la distribución condicional de probabilidades [1] $P_\xi(k \mid \eta)$, $k = 0, \ldots, N$ de la magnitud ξ para cualesquiera valores fijos $\eta(t)$, $t \in T$, *admitamos que* $\hat{\xi} = \varphi_0(\eta)$ *es igual al valor de* $k = 0, 1, \ldots, N$ *para el cual la probabilidad* $P_\xi(k \mid \eta)$ *es máxima*:

$$P_\xi(\hat{\xi} \mid \eta) = \max_{0 \leq k \leq N} P_\xi(k \mid \eta). \tag{3.11}$$

Teorema. *La magnitud* $\hat{\xi} = \varphi_0(\eta)$ *que satisface la condición* (3.11) *es la valuación óptima de* ξ.

Demostración. Introduzcamos las funciones $\chi_k(x)$, $k = 0, 1, \ldots, N$ que son iguales a 1 para $x = k$ e iguales a cero para los restantes valores de la variable x. Tomemos cualquiera magnitud $\varphi = \varphi(\eta)$ que tome uno de los valores $0, 1, \ldots, N$. Se ve fácilmente que

$$\mathbf{P}\{\varphi(\eta) \neq \xi\} = \mathbf{M}\zeta = \mathbf{M}[\mathbf{M}(\zeta \mid \eta)],$$

[1] La probabilidad $P_\xi(k \mid \eta)$ algunas veces la llaman *a posteriori* (después de la «observación» η).

donde la magnitud $\zeta = \frac{1}{2}\sum_{k=0}^{N} [\chi_k(\xi) - \chi_k(\varphi)]^2$ es igual a 1 para $\varphi(\eta) \neq \xi$ y es igual a cero en los casos restantes (para $\varphi(\eta) = \xi$). Para los valores fijos $\eta(t)$, $t \in T$, la magnitud $\chi_k(\varphi)$ es constante y $\mathbf{M}\{\chi_k(\xi) \mid \eta\} = P_\xi(k \mid \eta)$ de modo que

$$\mathbf{M}(\zeta \mid \eta) = \frac{1}{2}\sum_{k=0}^{N} \mathbf{M}\{[\chi_k(\xi) - \chi_k(\varphi)]^2 \mid \eta\} =$$

$$= \frac{1}{2}\sum_{k=0}^{N} \mathbf{M}\{[\chi_k(\xi) - P_\xi(k \mid \eta)]^2 \mid \eta\} + \frac{1}{2}\sum_{k=0}^{N} [P_\xi(k \mid \eta) - \chi_k(\varphi)]^2,$$

donde para $\varphi(\eta) = j$

$$\frac{1}{2}\sum_{k=0}^{N} [P_\xi(k \mid \eta) - \chi_k(\varphi)]^2 = \frac{1}{2}\sum_{k=0}^{N} P_\xi(k \mid \eta)^2 - P_\xi(j \mid \eta) + \frac{1}{2}.$$

Supongamos que $\zeta_0 = \frac{1}{2}\sum_{k=0}^{N} [\chi_k(\xi) - \chi_k(\varphi_0)]^2$ y examinemos la diferencia

$$\mathbf{M}(\zeta \mid \eta) - \mathbf{M}(\zeta_0 \mid \eta) = P_\xi(j_0 \mid \eta) - P_\xi(j \mid \eta),$$

donde $j_0 = \varphi_0(\eta)$ es aquel valor para el cual la probabilidad $P_\xi(j \mid \eta)$ es máxima (véase (3.11)). Está claro que para cualquier $j = \varphi(\eta)$

$$\mathbf{M}(\zeta \mid \eta) - \mathbf{M}(\zeta_0 \mid \eta) = \mathbf{M}\{\zeta - \zeta_0 \mid \eta\} \geqslant 0$$

y por consiguiente,

$$\mathbf{M}\zeta - \mathbf{M}\zeta_0 = \mathbf{M}(\zeta - \zeta_0) = \mathbf{M}[\mathbf{M}\{\zeta - \zeta_0 \mid \eta\}] \geqslant 0,$$

es decir

$$\mathbf{P}\{\varphi(\eta) \neq \xi\} \geqslant \mathbf{P}\{\varphi_0(\eta) \neq \xi\},$$

que es lo que se quería demostrar.

Ejemplo (*tarea sobre el desajuste*). Supongamos que nos interesa el estado del sistema en el momento variable de tiempo t, sobre el cual se podría juzgar sin error según la señal determinada $\theta(t)$, que caracteriza el paso de un estado a otro, pero de hecho «observamos» sólo el proceso aleatorio $\xi = \xi(t)$, $t \geqslant 0$, de tal modo que

$$d\xi(t) = \theta(t)\,dt + d\eta(t), \tag{3.12}$$

donde en la diferencial estocástica $d\xi(t)$ figura, como corrientemente, el proceso del movimiento browniano standard $\eta(t)$. Se exige dar la valuación del estado del sistema en el momento dado de tiempo t

según la trayectoria $\xi(s)$, $s \leqslant t$, más exactamente, dar la valuación de la magnitud $\theta(t)$.

Supongamos para mayor sencillez que el sistema se encuentre en el momento inicial en el estado 0 y en el transcurso del tiempo puede pasar a algún otro estado 1 (el paso puede significar, por ejemplo, el desajuste o rotura del sistema examinado).

Consideraremos que

$$\theta(t) = \begin{cases} 0 \text{ para } t < \tau, \\ 1 \text{ para } t \geqslant \tau, \end{cases}$$

donde el momento τ del paso de 0 a 1, es una magnitud aleatoria con la distribución exponencial de probabilidades:

$$\mathbf{P}\{\tau > t\} = e^{-\lambda t}.$$

Como fue indicado anteriormente, la valuación óptima del valor desconocido $\theta(t)$ por las magnitudes $\xi(s)$, $s \leqslant t$ (cuyo conjunto designemos, para abreviar, por ξ_t), se halla con ayuda de *las probabilidades a posteriori* [1])

$$\pi_0(t) = \pi(t), \quad \pi_1(t) = 1 - \pi(t),$$

donde

$$\pi(t) = \mathbf{P}\{\theta(t) = 0 \mid \xi_t\}.$$

Supongamos que el proceso aleatorio $\xi = \xi(t)$ se observa en los momentos discretos de tiempo t, múltiplos de Δt, y sigamos la evolución de las probabilidades a posteriori $\pi(t)$ en el transcurso del tiempo t.

Examinemos preliminarmente la densidad condicional de probabilidades $p\{x \mid \theta(t) = 0, \xi_t\}$ de la magnitud $\xi(t + \Delta t)$ para la condición $\theta(t) = 0$ y para la trayectoria fija $\xi_t = \{\xi(s), s \leqslant t\}$. Como se ve de la fórmula (3.12)

$$\xi(t + \Delta t) = \xi(t) + \int_t^{t+\Delta t} \theta(s)\, ds + \Delta\eta(t),$$

donde $\Delta\eta(t)$ es una magnitud de Gauss independiente con la media nula y dispersión Δt. Para la función fija $\theta(s)$, $t \leqslant s \leqslant t + \Delta t$, la densidad buscada es

$$\frac{1}{\sqrt{2\pi\Delta t}} \exp\left\{-\frac{1}{2\Delta t}\left[x - \xi(t) - \int_t^{t+\Delta t} \theta(s)\, ds\right]^2\right\}, \qquad (3.13)$$

$$-\infty < x < \infty,$$

y ya que

$$\int_t^{t+\Delta t} \theta(s)\, ds = \begin{cases} \Delta t - u \text{ para } \tau - t = u, \ 0 \leqslant u \leqslant \Delta t, \\ 0 \quad \text{para } \tau - t > \Delta t \end{cases}$$

[1]) *A posteriori* (del latín), después de la prueba.

y la magnitud aleatoria $\tau - t$ para la condición $\theta(t) = 0$ está distribuida en la semirecta $u \geqslant 0$ con la densidad de probabilidad $\lambda e^{-\lambda u}$, entonces obtenemos como resultado que

$$p\{x \mid \theta(t) = 0,\ \xi_t\} =$$

$$= \int_0^{\Delta t} \frac{1}{\sqrt{2\pi\Delta t}} \exp\left\{-\frac{1}{2\Delta t}[x - \xi(t) - (\Delta t - u)]^2\right\} \lambda e^{-\lambda u}\, du +$$

$$+ \frac{1}{\sqrt{2\pi\Delta t}} \exp\left\{-\frac{1}{2\Delta t}[x - \xi(t)]^2\right\} e^{-\lambda\Delta t}.$$

Para $\theta(t) = 0$, $\theta(t + \Delta t) = 0$ (es decir, de hecho para $\theta(s) \equiv 0$, $t \leqslant s \leqslant t + \Delta t$) la expresión (3,13) nos da la densidad condicional de probabilidad

$$p\{x \mid \theta(t) = 0,\ \xi_t,\ \theta(t + \Delta t) = 0\} = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left\{-\frac{1}{2\Delta t}[x - \xi(t)]^2\right\}$$

Se comprueba fácilmente que tiene lugar la igualdad siguiente, análoga a la fórmula de Bayes (3.7):

$$\mathbf{P}\{\theta(t + \Delta t) = 0 \mid \theta(t) = 0,\ \xi_t,\ \xi(t + \Delta t)\} =$$

$$= \mathbf{P}\{\theta(t + \Delta t) = 0 \mid \theta(t) = 0,\ \xi_t\} \times$$

$$\times \frac{p\{\xi(t + \Delta t) \mid \theta(t) = 0,\ \xi_t,\ \theta(t + \Delta t) = 0\}}{p\{\xi(t + \Delta t) \mid \theta(t) = 0,\ \xi_t\}}, \qquad (3.14)$$

en la cual

$$\frac{p\{\xi(t + \Delta t) \mid \theta(t) = 0,\ \xi_t\}}{p\{\xi(t + \Delta t) \mid \theta(t) = 0,\ \xi_t,\ \theta(t + \Delta t) = 0\}} =$$

$$= \int_0^{\Delta t} \exp\left\{\frac{1}{2\Delta t}[2\Delta\xi(t)(\Delta t - u) - (\Delta t - u)^2]\right\} \lambda e^{-\lambda u}\, du +$$

$$+ e^{-\lambda\Delta t} = [1 + o(1)]\,\lambda\Delta t + e^{-\lambda\Delta t} = 1 + o(\Delta t),$$

donde la magnitud $o(\Delta t)$, dependiente del incremento $\Delta\xi(t) = \xi(t + \Delta t) - \xi(t)$, es una magnitud infinitesimal de orden superior en comparación con Δt para $\Delta t \to 0$ (cuando $\Delta\xi(t) \to 0$). Teniendo en cuenta que $\mathbf{P}\{\theta(t + \Delta t) = 0 \mid \theta(t) = 0,\ \xi_t\}$ coincide con la probabilidad de quedarse en el estado 0 durante un intervalo de tiempo Δt (esta probabilidad es igual a $e^{-\lambda\Delta t}$), de la igualdad (3.14) obtenemos la relación siguiente:

$$\mathbf{P}\{\theta(t + \Delta t) = 0 \mid \theta(t) = 0,\ \xi_{t+\Delta t}\} = 1 - \lambda\,\Delta t + o(\Delta t),$$

donde la magnitud $o(\Delta t)$ tiene el mismo sentido que anteriormente. Luego, análogamente a (3.14) tiene lugar la igualdad

$$\mathbf{P}\{\theta(t) = 0 \mid \xi_t,\ \xi(t + \Delta t)\} =$$

$$= \mathbf{P}\{\theta(t) = 0 \mid \xi_t\} \frac{p\{\xi(t + \Delta t) \mid \theta(t) = 0,\ \xi_t\}}{p\{\xi(t + \Delta t) \mid \xi_t\}}. \qquad (3.15)$$

Evidentemente,

$$p\{\xi(t+\Delta t) \mid \xi_t\} = p\{\xi(t+\Delta t) \mid \theta(t) = 0, \xi_t\}\pi_0(t) + p\{\xi(t+\Delta t) \mid \theta(t) = 1, \xi_t\}\pi_1(t),$$

donde la condición $\theta(t) = 1$ de hecho significa que $\theta(s) = 1$ para todos los $s \geqslant t$; de la fórmula general (3.13) obtenemos

$$\frac{p\{\xi(t+\Delta t) \mid \theta(t)=1, \xi_t\}}{p\{\xi(t+\Delta t) \mid \theta(t)=0, \xi_t\}} = \exp\left\{\Delta\xi(t) - \frac{\Delta t}{2}\right\}(1+o(\Delta t)),$$

y la igualdad (3.15) permite establecer que

$$\mathbf{P}\{\theta(t)=0 \mid \xi_{t+\Delta t}\} = \mathbf{P}\{\theta(t)=0 \mid \xi_t, \xi(t+\Delta t)\} =$$

$$= \pi_0(t)\frac{1}{\pi_0(t)+\pi_1(t)+\exp\left\{\Delta\xi(t)-\frac{\Delta t}{2}\right\}(1+o(\Delta t))} =$$

$$= \pi_0(t)\frac{1}{1+\pi_1(t)\left(\Delta\xi(t)-\frac{\Delta t}{2}+\frac{1}{2}\Delta\xi(t)^2\right)+o(\Delta t)} =$$

$$= \pi_0(t)\left\{1-\pi_1(t)\left(\Delta\xi(t)-\frac{\Delta t}{2}+\frac{1}{2}\Delta\xi(t)^2\right)+\right.$$

$$\left.+\left[\pi_1(t)\left(\Delta\xi(t)-\frac{\Delta t}{2}+\frac{1}{2}\Delta\xi(t)^2\right)\right]^2+o(\Delta t)\right\} =$$

$$= \pi_0(t)\left\{1-\pi_1(t)\left(\Delta\xi(t)-\frac{\Delta t}{2}+\frac{1}{2}\Delta\xi(t)^2\right)+\pi_1(t)^2\Delta\xi(t)^2+o(\Delta t)\right\},$$

donde $o(\Delta t)$ ahora significa una magnitud de un orden infinitesimal superior en comparación con $\Delta t + \Delta\xi(t)^2$.

Por último, utilizando la fórmula

$$\mathbf{P}\{\theta(t+\Delta t)=0 \mid \xi_{t+\Delta t}\} = \mathbf{P}\{\theta(t+\Delta t)=0 \mid \theta(t)=0, \xi_{t+\Delta t}\}\,\mathbf{P}\{\theta(t)=0 \mid \xi_{t+\Delta t}\},$$

de las relaciones obtenidas anteriormente, obtenemos para $\pi(t) = \pi_0(t) = 1 - \pi_1(t)$:

$$\pi(t+\Delta t) = (1-\lambda\Delta t)\pi(t)\left[1-(1-\pi(t))\left(\Delta\xi(t)-\frac{\Delta t}{2}+\frac{1}{2}\Delta\xi(t)^2\right)+\right.$$

$$\left.+(1-\pi(t))^2\Delta\xi(t)^2\right]+o(\Delta t) = \pi(t)-\pi(t)\times$$

$$\times\left\{\lambda\Delta t+\frac{1}{2}(1-\pi(t))[\Delta\xi(t)^2-\Delta t]-(1-\pi(t))^2\Delta\xi(t)^2\right\}+$$

$$+\pi(t)(1-\pi(t))\Delta\xi(t)+o(\Delta t).$$

Si despreciamos aquí la magnitud $o(\Delta t)$, que tiene un orden superior en comparación con $\Delta t + \Delta\xi(t)^2$, y también el miembro $\frac{1}{2}\times$ $\times(1-\pi(t))[\Delta\xi(t))^2 - \Delta t]$, que contiene la magnitud $\Delta\xi(t)^2 - \Delta t$

para la cual

$$\mathbf{M}\{\Delta\xi(t)^2 - \Delta t \mid \xi_t\} = o(\Delta t)$$

y

$$\mathbf{M}\{[\Delta\xi(t)^2 - \Delta t]^2 \mid \xi_t\} = o(\Delta t)$$

(compárese con las condiciones (8.13), (8.14)), entonces obtenemos

$$\Delta\pi(t) \sim -\pi(t)(\lambda - (1 - \pi(t)^2)\Delta t + \pi(t)(1 - \pi(t))\Delta\xi(t).$$

Examinando el caso cuando t es continua se puede indicar que $\pi(t) = \mathbf{P}\{\theta(t) = 0 \mid \xi_t\}$ como función del proceso aleatorio $\xi(s)$, $s \leqslant t$, varía continuamente con el transcurso del tiempo t de tal modo que existe la diferencial estocástica

$$d\pi(t) = -\pi(t)(\lambda - (1 - \pi(t)^2)\,dt + \pi(t)(1 - \pi(t))\,d\xi(t),$$

donde la diferencial estocástica $d\xi(t)$ está determinada por la fórmula (3.12).

INDICE DE MATERIAS

A

Adición de probabilidades 19
Aditividad 19
— numerable 20
Aproximación de Poissón para la distribución binomial 96
— óptima 253

C

Cadena de Markov 141
— de Markov de tiempo contínuo 161
Característica espectral del sistema lineal 222
Clase cerrada de estados 153
Coeficiente de correlación 58
— de difusión 102
— de ergodicidad 154, 167
Combinaciones 73
Composición de las densidades de probabilidades 41
Complejo de aparatos 228
Condición de Lipshitz 241
— de Liapunov 114, 134
Continuidad de la probabilidad 20
Control por elección de la producción 80
Convergencia hacia la distsibución de probabilidades 127
— hacia la distribución estacionaria 153, 167
Corriente de exigencias de Poissón 163
— de sucesos de Poissón 87, 138
— — — — — compleja 210

D

Degeneración del proceso que se ramifica 176
Densidad de distribución condicional 43
— — — conjunta 36
Densidad de estructura 203
— de paso del estado 162
— de paso del proceso de Markov 139
— de probabilidad 11, 12, 15, 35, 36
— de salida del estado 144, 162
— espectral del proceso estacionario 216
Derivada media cuadrática 227
Desigualdad de Cauchy—Bunjakovsky 51
— de Chevishev 51
— del triángulo 64
Desintegración radiactiva 87, 90, 137
Desviación standard 109
Diferencia de sucesos 17
Diferencial estocástica 240
Difusión 232
Dispersión 52
— de la magnitud compleja 202
— empírica 116
Distancia media cuadrática 57
Distribución de las partículas 76, 84
— de las partículas de Poissón 86
— de Maxwell 117
— de probabilidades 33
— — — binomial (distribución de Bernoulli) 82, 96
— — — de Cauchy 124
— — — condicional 42
— — — conjunta 36, 44
— — — conjunta continua 36
— — — *continua* 34
— — — de Gauss 20
— — — de Maxwell 117
— — — de Poissón 81, 83, 86
— — — de Rayleigh 117
Distribución de probabilidades de Student 119, 124
— — — discreta 34
— — — en triángulo 42
— — — exponencial 89
— — — geométrica 89
— — — *hipergeométrica* 80
— — — normal (de Gauss) 98, 109, 111, 201
— — — normal doble 106
— — — normal standard 109
— del proceso aleatorio de dimensiones finitas 135
Distribución de Student 119, 124
— — gamma 93, 117
— de X—cuadrado 117
— *t* 119

E

Ecuación de Kolmogorov 162, 248
— de Kolmogorov-Chapmen 245
Elección con retorno 72, 81
Efecto fraccionario 209
Energía media del proceso estacionario 215
Envoltura de las densidades de probabilidades 41
Espacio de sucesos elementales 16
— lineal 253
Esperanza matemática 45, 48
— — condicional 53, 271
— — repetida 55
— — total 55
Espectro de frecuencias del proceso estacionario 215
Estado accesible 152
— absorvente 144
— irreversible 145
Estado positivo 152
— reversible 145
Estados comunicantes 152

F

Fenómeno de explosión 177
Fluctuación aleatoria 137, 147, 159
Fórmula de Bayes 275
— de Erlang 169
— de inversión 122, 228
— de la esperanza matemática total 55
— de la probabilidad total 31, 56
— de Stirling 75
Frecuencia del proceso estacionario 215
— del suceso 70
Función característica 121
— de correlación 135
— de correlación del proceso estacionario 214
— de distribución 34
— de peso 199, 207
— de transmisión del sistema lineal 222
— seleccionada del proceso aleatorio 136
— — gamma de Euler 94
— productora 146

H

Hiperplano 253

I

Identidad de Plancherel 132
— de Wald 195
Igualdad de Párceval 131
Independencia de las magnitudes aleatorias 45
— de las pruebas 25
Indicador de sucesos 56
Integración 225
Integral 226
— de Fourier 121
— de probabilidad 108
— estocástica 203, 205, 234
Intersección (multiplicación) de sucesos 17

J

Juego a la ruleta 10, 11
— de preference 76
— hasta la primera pérdida 21

L

Lanzamiento del dado de juego 9, 18
— de la moneda 9, 21, 88
Lema de Borel—Cantelli 63
— sobre la perpendicular 253
Ley del arco seno 108
— de distribución (véase la distribución de probabilidades)
— de la suma de probabilidades 19
— de los números grandes 69, 270
— reforzada de los números grandes 69

M

Matriz de correlación 69, 110
— determinada positivamente 60
— de correlación 60, 110
Magnitudes aleatorias independientes 84
Magnitud aleatoria 11
— discreta 34
— distribuida uniformemente 12
Momento de la magnitud aleatoria 51, 122
Momento de tiempo de Markov 140
Movimiento browniano 100, 138, 233
— perturbado del péndulo 231
Multiplicación pura 166

N

Número de combinaciones 73
— de grados de libertad 117
— de permutaciones 73
— de variaciones 73

P

Partícula browniana y demás 100
Período de semidesintegración 91
Permutación 73
Perpendicular al subespacio 253
Probabilidad 10, 16, 71
— a posteriori 277, 279
— condicional 29, 56
— de explosión 177
— del nacimiento de un niño 23
— de paso 139, 141
— de reversibilidad 145
Proceso aleatorio 135
— — con incrementos no correlacionados 202
— — complejo 202
— — complejo de Poissón 210
— — de difusión de Markov 243
— — de efecto fraccionario 209
— — de Gauss 135
— — de Markov 136
— — de Poissón 138
Proceso de restablecimiento 177
— aleatorio estacionario en ámplio sentido 208, 214
Proceso aleatorio estacionario con espectro continuo 215
— — estacionario con espectro discreto 214
— — homogéneo con incrementos independientes 209
Proceso aleatorio homogéneo de Markov 139
— — que se ramifica de Markov 170
— — riguroso de Markov 140
— puro de multiplicación 165
Proyección en el hiperplano 253
Prueba con un número finito de resultados equiprobables 10
— de Bernoulli 95
Pruebas independientes 29

R

Ruido blanco 231

S

Sistema con tiempo exponencial de servicio 164
— desarrollado de sucesos 23
— de servicio 168, 177
— de servicio con una línea 191
— directo de ecuaciones diferenciales de Kolmogorov 162
— inverso de ecuaciones diferenciales de Kolmogorov 162
— lineal 199
— lineal estable 207
— completo de sucesos 31
Sucesión seleccionada 61
Suceso cierto 17
— complementario 17
— elemental 10, 16
— imposible 17
Sucesos incompatibles (que no se cortan) 17, 18
— independientes 28
Suma de sucesos 17
— integral 225

T

Tablero de Galtón 103
Tarea del complejo de aparatos 255
— sobre la aguja de Buffón 13
— sobre el arruinamiento del jugador 31
— sobre el complejo de aparatos 255
— sobre el desajuste 278
— sobre la distribución de las partículas 76, 84
Teorema central límite 114, 133
— de Moivre Laplace 97, 114
— ergódico 153
Transformación de Laplace 222
— de Fourier 220
— inversa 121
Trayectoria del proceso aleatorio 135

U

Unión de sucesos 17

V

Valor condicional 53
— empírico 115
— medio 45, 48, 135
— de la media cuadrática 51
Valuaciones de los cuadrados mínimos 266
— lineales óptimas 266
— no desplazadas 264
Variaciones 73

A NUESTROS LECTORES:

« MIR » edita libros soviéticos traducidos al español, inglés, francés y árabe. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica; manuales para los centros de enseñanza superior y escuelas tecnológicas; literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia ficción.

Dirijan sus opiniones a Editorial « MIR » I Rizhski per. 2, GSP I-110, Moscú, 129820, URSS.

EN 1974 LA EDITORIAL MIR PUBLICARA:

GMURMAN V.

La teoría de las probabilidades y la estadística matemática

Este tratado ha sido escrito de conformidad con el nuevo curso de la teoría de las probabilidades y de la estadística matemática. El material de esta obra se presenta en tres partes fundamentales: las dos primeras se dedican a la teoría de las probabilidades y la última, a la estadística matemática. En este manual se estudian los temas siguientes: probabilidad de las hipótesis; fórmulas de Bayes; distribución de Poisson; nociones de la estadística matemática y noción de correlación.

En el libro se presta gran atención a los métodos estadísticos de elaboración de los datos experimentales; las tablas que se exponen para los cálculos son muy cómodas. Cada capítulo contiene problemas y sus respuestas que han sido elegidos adecuadamente. Además, todo capítulo va acompañado del análisis de las soluciones de los problemas del material correspondiente.

Esta obra contiene diecisiete capítulos, varias tablas de números, veintidós figuras y un gran número de ejemplos teóricos y técnicos; se recomienda para los estudiantes de las facultades ingeniero—técnicas y económicas.

En 1975 se editará:

GMURMAN V.

Obra de referencia para resolver los problemas de la teoría de las probabilidades y de la estadística matemática

Por su contenido, esta obra corresponde al curso de la teoría de las probabilidades y de estadística matemática y contiene más de 500 problemas para todas las partes del curso. Al principio de cada párrafo el autor expone los conocimientos teóricos necesarios, las fórmulas requeridas y muestra los ejemplos cómo resolver problemas típicos. Los problemas destinados para resolverse independientemente se dan según sus dificultades crecientes. Además de los problemas clásicos acerca de lanzamientos de monedas o dados, este libro contiene un gran número de problemas que ilustran la aplicación de la teoría de las probabilidades en la técnica (fiabilidad, control estadístico de la calidad de producción y tratamiento de los datos experimentales). El mérito fundamental de este libro consiste en que el autor, en todos los problemas, pone en primer plano su esencia probabilística, reduciendo al mínimo el cálculo numérico.

Esta obra puede servir para enseñar a resolver los problemas de la teoría de las probabilidades, de estadística matemática y para los estudiantes de las especialidades ingenierotécnicas y económicas.